

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 78.4648 Seconds  
(without alignments)  
44.362 Million cell updates/sec

Title: US-10-067-484-2  
Perfect score: 43  
Sequence: 1 PTSFNXATK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	41	95.3	9	5	ABB81969	Abb81969 30 kDa ra
2	34	79.1	424	6	ABU41254	Abu41254 Protein e
3	33	76.7	135	3	AAB24649	Aab24649 Plant SDF
4	33	76.7	271	3	AAG10010	Aag10010 Arabidops
5	33	76.7	345	4	ABB60198	Abb60198 Drosophil
6	32	74.4	232	7	ABO60780	Abo60780 Klebsiell
7	32	74.4	360	4	AAM41900	Aam41900 Human pol
8	32	74.4	663	4	ABB52462	Abb52462 Escherich
9	32	74.4	713	7	ADC01362	Adc01362 Enterohae

10	32	74.4	808	7	ADJ68916	Adj68916	Human hea
11	32	74.4	2194	4	AAM40114	Aam40114	Human pol
12	32	74.4	2829	8	ADL72180	Adl72180	X. laevis
13	31	72.1	110	4	AAB62783	Aab62783	Human HIV
14	31	72.1	126	4	ABG02725	Abg02725	Novel hum
15	31	72.1	457	4	ABB69700	Abb69700	Drosophil
16	31	72.1	1361	6	ABU20358	Abu20358	Protein e
17	31	72.1	6815	4	ABB66811	Abb66811	Drosophil
18	30	69.8	85	1	AAP70412	Aap70412	ORF 9 gen
19	30	69.8	100	2	AAY35084	Aay35084	Chlamydia
20	30	69.8	110	6	ABP60080	Abp60080	Mouse RGS
21	30	69.8	154	7	ABO65493	Abo65493	Klebsiell
22	30	69.8	187	4	ABG04966	Abg04966	Novel hum
23	30	69.8	214	7	ADH86644	Adh86644	Enterococ
24	30	69.8	287	7	ABO74345	Abo74345	Pseudomon
25	30	69.8	319	4	AAU63426	Aau63426	Propionib
26	30	69.8	319	6	ABM65512	Abm65512	Propionib
27	30	69.8	319	6	ABM59945	Abm59945	Propionib
28	30	69.8	371	8	ADN24100	Adn24100	Bacterial
29	30	69.8	407	6	ABU37853	Abu37853	Protein e
30	30	69.8	418	5	ABB06969	Abb06969	Canine AS
31	30	69.8	426	6	ABP77795	Abp77795	N. gonorr
32	30	69.8	426	6	ABU37534	Abu37534	Protein e
33	30	69.8	441	6	ADA35082	Ada35082	Acinetoba
34	30	69.8	851	8	ADQ66949	Adq66949	Novel hum
35	30	69.8	870	2	AAR31348	Aar31348	Jaagsiekt
36	30	69.8	987	8	ADL05331	Adl05331	M. catarr
37	30	69.8	1362	6	ABU17069	Abu17069	Protein e
38	30	69.8	1363	6	ABU35175	Abu35175	Protein e
39	30	69.8	1380	6	ADA36845	Ada36845	Acinetoba
40	30	69.8	2626	3	AAB07569	Aab07569	Protein e
41	29	67.4	65	4	AAU45762	Aau45762	Propionib
42	29	67.4	65	6	ABM42281	Abm42281	Propionib
43	29	67.4	106	2	AAW71241	Aaw71241	Light cha
44	29	67.4	112	7	ADD13784	Add13784	vKappa ex
45	29	67.4	127	2	AAW08945	Aaw08945	Kappa lig

# ALIGNMENTS

## RESULT 1

ABB81969

ID ABB81969 standard; peptide; 9 AA.

XX

AC ABB81969;

XX

DT 25-NOV-2002 (first entry)

XX

DE 30 kDa ragweed pollen allergen tryptic peptide 2.

XX

KW Ragweed; pollen; allergen; Ambt 7; glycoprotein; antiallergic;  
KW immunotherapy; disulphide protein.

XX

OS Ambrosia elatior.

XX

FH Key Location/Qualifiers

FT Misc-difference 6  
 FT /label= Leu or Ile  
 XX  
 PN WO200263012-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 04-FEB-2002; 2002WO-US003346.  
 XX  
 PR 05-FEB-2001; 2001US-0266686P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Buchanan BB, Del Val G, Frick OL;  
 XX  
 DR WPI; 2002-657539/70.  
 XX  
 PT New ragweed pollen allergens, useful in allergy testing and immunotherapy  
 PT regimens, particularly for treating sensitivity to pollen or pollen  
 PT allergy (e.g. anaphylaxis, or symptoms of hives or asthma) in a mammal,  
 PT especially a human.  
 XX  
 PS Claim 1; Page 53; 70pp; English.  
 XX  
 CC The invention relates to an isolated pollen allergen purified from  
 CC ragweed pollen, substantially free of any other pollen proteins, or a  
 CC protein that is an antigenic fragment of a pollen allergen Ambt 7. The  
 CC allergen is characterized by the following physiochemical and biological  
 CC properties: (a) being contained in pollen extracts; (b) a glycoprotein;  
 CC (c) a sulphhydryl group containing protein; (d) a molecular weight of  
 CC about 30 kDa as determined by SDS-polyacrylamide gel electrophoresis; and  
 CC (e) possessing allergen activity. The pollen allergen, or antigenic  
 CC protein fragment of the pollen allergen Ambt 7, or composition is useful  
 CC for treating sensitivity to pollen or pollen allergy in a mammal. This  
 CC allergy includes anaphylaxis or atopy, which includes the symptoms of hay  
 CC fever, asthma or hives. The allergen is also useful in allergy testing  
 CC and immunotherapy regimens. Sequences ABB81968-978 represent tryptic  
 CC peptide fragments of the 30 kDa ragweed complete pollen extract  
 CC disulphide protein allergen  
 XX  
 SQ Sequence 9 AA;

Query Match 95.3%; Score 41; DB 5; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9  
 |||||  
 Db 1 PTSFNXATK 9

RESULT 2  
 ABU41254  
 ID ABU41254 standard; protein; 424 AA.  
 XX  
 AC ABU41254;  
 XX

DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #26781.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS *Proteus* sp.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA45124.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 69178; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 424 AA;

Query Match 79.1%; Score 34; DB 6; Length 424;  
Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||| | |:  
Db 368 PTSFNSVTE 376

### RESULT 3

AAB24649

ID AAB24649 standard; peptide; 135 AA.

XX

AC AAB24649;

XX

DT 06-AUG-2003 (revised)

DT 27-NOV-2000 (first entry)

XX

DE Plant SDF encoded polypeptide sequence SEQ List 1 NO:63.

XX

KW Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment; SDF;  
KW genetic mapping; identification; promoter; structural gene; UTR;  
KW untranslated region; expression control.

XX

OS Viridiplantae.

XX

PN WO200040695-A2.

XX

PD 13-JUL-2000.

XX

PF 07-JAN-2000; 2000WO-US000466.

XX

PR 08-JAN-1999; 99US-0115293P.

XX

PA (CERE-) CERES INC.

XX

PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;

PI Zheng L;

XX

DR WPI; 2000-465970/40.

XX

PT New corn plant and Arabidopsis thaliana sequence-determined DNA  
PT fragments, useful for expressing gene products and for controlling  
PT expression of a target gene.

XX

PS Claim 14; Page 355; 673pp; English.

XX  
 CC The present invention describes polynucleotides, such as complete cDNA  
 CC sequences and/or sequences of genomic DNA encompassing complete genes,  
 CC portions of genes, and/or intergenic regions, collectively referred to as  
 CC sequence-determined DNA fragments (SDFs), from corn plants and  
 CC Arabidopsis thaliana. The SDFs are promoters, structural genes,  
 CC untranslated regions (UTRs), or 3' termination sequences. They can be  
 CC used for expressing a gene product and controlling expression of a target  
 CC gene, either as a promoter, a structural gene, an UTR or as a 3'  
 CC termination sequence. They are also useful as tools for genetic mapping,  
 CC and identification of a particular individual plant or for clustering a  
 CC group of plants with a common trait. AAA78433 to AAA78630 and AAB24605 to  
 CC AAB25099 represent the specifically claimed polynucleotide sequences and  
 CC polypeptides encoded by them given in the present invention. (Updated on  
 CC 06-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 135 AA;

Query Match 76.7%; Score 33; DB 3; Length 135;  
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||:|: |||  
 Db 16 PTTFSVATK 24

#### RESULT 4

AAG10010

ID AAG10010 standard; protein; 271 AA.  
 XX  
 AC AAG10010;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 8162.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-00301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127462P.  
 PR 06-APR-1999; 99US-0128234P.

PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	16-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139492P.
PR	18-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139460P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	22-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	23-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140991P.

PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	01-JUL-1999;	99US-0142154P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.



PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.

PR 29-OCT-1999; 99US-0162142P.

Query Match 76.7%; Score 33; DB 3; Length 271;  
Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9  
||:|: |||  
Db 30 PTTFSVATK 38

RESULT 5

ABB60198

ID ABB60198 standard; protein; 345 AA.

XX

AC ABB60198;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 7386.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL04301.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signaling and cell-cell  
interactions.

XX

PS Disclosure; SEQ ID NO 7386; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 345 AA;

Query Match 76.7%; Score 33; DB 4; Length 345;

Best Local Similarity 66.7%; Pred. No. 3.1e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9

||: |||

Db 25 PTTINSATK 33

# RESULT 6

ABO60780

ID ABO60780 standard; protein; 232 AA.

XX

AC ABO60780;

XX

DT 29-JUL-2004 (first entry)

XX

DE Klebsiella pneumoniae polypeptide seqid 7297.

XX

KW Recombinant expression vector; transcription regulatory element;

KW Klebsiella pneumoniae protein; antibacterial; Vaccine.

XX

OS Klebsiella pneumoniae.

XX

PN US6610836-B1.

XX

PD 26-AUG-2003.

XX

PF 27-JAN-2000; 2000US-00489039.

XX

PR 29-JAN-1999; 99US-0117747P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Breton GL, Osborne M;

XX

DR WPI; 2003-895346/82.

DR N-PSDB; ACH94331.

XX

PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.

XX

PS Disclosure; SEQ ID NO 7297; 932pp; English.

XX

CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This is the amino acid sequence of a  
CC Klebsiella pneumoniae polypeptide of the invention

XX

SQ Sequence 232 AA;

Query Match 74.4%; Score 32; DB 7; Length 232;  
Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
| ||| ||:  
Db 157 PRSFNAATE 165

RESULT 7

AAM41900

ID AAM41900 standard; protein; 360 AA.

XX

AC AAM41900;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 6831.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US034263.

XX

PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

PI Zhou P, Goodrich R, Drmanac RT;

XX

DR WPI; 2001-442253/47.

DR N-PSDB; AAI61056.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.

XX

PS Example 2; SEQ ID NO 6831; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification

XX

SQ Sequence 360 AA;

Query Match 74.4%; Score 32; DB 4; Length 360;

Best Local Similarity 66.7%; Pred. No: 5.1e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9

||:|| | |

Db 41 PTNFNVAEK 49

#### RESULT 8

ABB52462

ID ABB52462 standard; protein; 663 AA.

XX

AC ABB52462;

XX

DT 11-FEB-2002 (first entry)

XX

DE Escherichia coli polypeptide SEQ ID NO 263.

XX

KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
KW systemic infection; non-diarrhoeal infection; septicaemia;  
KW pyelonephritis; antibiotic resistance.

XX

OS Escherichia coli.

XX

PN WO200166572-A2.

XX

PD 13-SEP-2001.

XX

PF 12-MAR-2001; 2001WO-EP003445.

XX

PR 10-MAR-2000; 2000FR-00003145.

PR 02-FEB-2001; 2001FR-00001449.

XX

PA (INRM ) INSERM INST. NAT SANTE & RECH MEDICALE.

XX

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;  
 XX  
 DR WPI; 2001-550253/61.  
 XX  
 PT A library of DNA fragments of Escherichia coli strains for the phylogenic  
 PT determination of a given strain comprises polynucleotides of nature B2/D+  
 PT A-.  
 XX  
 PS Example 6; Fig 6; 646pp; English.  
 XX  
 CC The invention relates to a library of DNA fragments of Escherichia coli  
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and  
 CC encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature  
 CC B2/D+A-. The polynucleotides have potential antiinflammatory,  
 CC antibacterial and immunosuppressive activity as part of pharmaceutical  
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
 CC infections. The polypeptides are useful for determining the phylogenic  
 CC group of a given E. coli strain. These polypeptides can detect and treat  
 CC an undesired development of E. coli, particularly an extra-intestinal  
 CC infection that include systemic and non-diarrhoeal infections such as  
 CC septicaemia, pyelonephritis and meningitis this is particularly  
 CC advantageous as bacterial resistance is increasing with the more frequent  
 CC use of broad spectrum antibiotics  
 XX  
 SQ Sequence 663 AA;

Query Match 74.4%; Score 32; DB 4; Length 663;  
 Best Local Similarity 66.7%; Pred. No. 9.6e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 | | | | |  
 Db 140 PRSFNAATE 148

RESULT 9  
 ADC01362

ID ADC01362 standard; protein; 713 AA.  
 XX  
 AC ADC01362;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 1407.  
 XX  
 KW enterohaemorrhagic; anti-bacterial.  
 XX  
 OS Escherichia coli; 0157:H7.  
 XX  
 PN JP2002355074-A.  
 XX  
 PD 10-DEC-2002.  
 XX  
 PF 24-JAN-2002; 2002JP-00015959.  
 XX  
 PR 24-JAN-2001; 2001JP-00112010.  
 XX

PA (UYTS-) UNIV TSUKUBA.  
 XX  
 DR WPI; 2003-451640/43.  
 XX  
 PT Enterohemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule  
 PT and a polypeptide and its use, a polypeptide, a vector and a host cell.  
 XX  
 PS Claim 3; SEQ ID NO 1407; 2067pp; Japanese.  
 XX  
 CC The invention relates to a novel enterohaemorrhagic Escherichia coli  
 CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention  
 CC has anti-bacterial activity. The polypeptide can be used in detection  
 CC and/or treatment of O157:H7 infection. The nucleotide sequence of the  
 CC genome of Enterohaemorrhagic E coli O157:H7 was determined. The present  
 CC sequence represents an E. coli O157:H7-specific polypeptide of the  
 CC invention.  
 XX  
 SQ Sequence 713 AA;

Query Match 74.4%; Score 32; DB 7; Length 713;  
 Best Local Similarity 66.7%; Pred. No. 1e+03;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 | | | | |  
 Db 140 PRSFNAATE 148

# RESULT 10

ADJ68916

ID ADJ68916 standard; protein; 808 AA.  
 XX  
 AC ADJ68916;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human heat mitochondrial protein as a therapeutic target SeqID722.  
 XX  
 KW mitochondrial; human; screening assay; diabetes mellitus;  
 KW Huntington's disease; osteoarthritis;  
 KW Leber's hereditary optic neuropathy; LHON;  
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
 KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;  
 KW osteopathic; ophthalmological; cytostatic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003087768-A2.  
 XX  
 PD 23-OCT-2003.  
 XX  
 PF 04-APR-2003; 2003WO-US010870.  
 XX  
 PR 12-APR-2002; 2002US-0372843P.  
 PR 17-JUN-2002; 2002US-0389987P.  
 PR 20-SEP-2002; 2002US-0412418P.

XX  
PA (MITO-) MITOKOR.  
PA (BUCK-) BUCK INST AGE RES.  
XX  
PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;  
PI Warnock DE;  
XX  
DR WPI; 2003-845369/78.  
XX  
PT Identifying a mitochondrial target for drug screening assays and for  
PT treating diseases associated with altered mitochondrial function,  
PT comprises detecting a modified polypeptide in a sample and correlating  
PT with the disease.  
XX  
PS Claim 1; SEQ ID NO 722; 180pp; English.  
XX  
CC This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nootropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.  
XX  
SQ Sequence 808 AA;

Query Match 74.4%; Score 32; DB 7; Length 808;  
Best Local Similarity 66.7%; Pred. No. 1.2e+03;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||:||||  
Db 489 PTNFNVAEK 497

# RESULT 11

AAM40114

ID AAM40114 standard; protein; 2194 AA.

XX

AC AAM40114;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 3259.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;



KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US034263.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AAI59270.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Example 5; SEQ ID NO 3259; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 XX  
 SQ Sequence 2194 AA;

Query Match 74.4%; Score 32; DB 4; Length 2194;  
 Best Local Similarity 66.7%; Pred. No. 3.4e+03;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||: || |  
Db 515 PTNFNVAEK 523

RESULT 12

ADL72180

ID ADL72180 standard; protein; 2829 AA.

XX

AC ADL72180;

XX

DT 20-MAY-2004 (first entry)

XX

DE X. laevis mutated adenomatous polyposis coli (APC) protein.

XX

KW APC; adenomatous polyposis coli; polyp; cancer; mutant.

XX

OS Xenopus laevis.

OS Synthetic.

XX

PN WO2004018677-A1.

XX

PD 04-MAR-2004.

XX

PF 19-AUG-2003; 2003WO-JP010434.

XX

PR 22-AUG-2002; 2002JP-00241487.

XX

PA (EISA ) EISAI CO LTD.

XX

PI Kiyosue Y, Sasaki H, Tsukita S;

XX

DR WPI; 2004-238977/22.

XX

PT Mutated adenomatous polyposis coli protein induces multi-layering of  
PT cells involved in polyp and cancer formation.

XX

PS Claim 2; SEQ ID NO 1; 68pp; Japanese.

XX

CC The invention relates to a mutated APC (adenomatous polyposis coli)  
CC protein that induces multi-layering of cells. The invention also provides  
CC a method for screening for compounds that inhibit multi-layering of  
CC cells. The protein can be used for investigating the mechanisms involved  
CC in polyp and cancer formation. The present sequence represents a mutated  
CC APC protein sequence.

XX

SQ Sequence 2829 AA;

Query Match 74.4%; Score 32; DB 8; Length 2829;  
Best Local Similarity 66.7%; Pred. No. 4.4e+03;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||||: ||  
Db 1927 PTSFSSAAK 1935

RESULT 13

AAB62783

ID AAB62783 standard; protein; 110 AA.  
 XX  
 AC AAB62783;  
 XX  
 DT 03-APR-2001 (first entry)  
 XX  
 DE Human HIV-1 monoclonal antibody SEQ ID NO: 82.  
 XX  
 KW Human immunodeficiency virus-1; HIV-1; human monoclonal antibody;  
 KW envelope glycoprotein; gp120; diagnosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200100678-A1.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-US017327.  
 XX  
 PR 30-JUN-1999; 99US-0141701P.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Watkins BA, Reitz MS;  
 XX  
 DR WPI; 2001-112438/12.  
 DR N-PGDB; AAF29084.  
 XX  
 PT Novel human monoclonal antibody immunoreactive with human  
 PT immunodeficiency virus-1 glycoprotein gp120, useful for detecting HIV-1  
 PT in biological sample and providing passive immunotherapy to HIV-1  
 PT infected mammal.  
 XX  
 PS Claim 1; Page 74; 81pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences for the  
 CC variable regions of human monoclonal antibodies which are immunoreactive  
 CC with human immunodeficiency virus-1 (HIV-1) envelope glycoprotein gp120.  
 CC These can be used in diagnosis and therapy of HIV-1 infection  
 XX  
 SQ Sequence 110 AA;

Query Match 72.1%; Score 31; DB 4; Length 110;  
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 |||| ||  
 Db 97 PTSFGQGTK 105

RESULT 14

ABG02725

ID ABG02725 standard; protein; 126 AA.  
 XX

AC ABG02725;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #2716.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS66912.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 33084; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 126 AA;

Query Match 72.1%; Score 31; DB 4; Length 126;  
Best Local Similarity 66.7%; Pred. No. 2.7e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||| |  
Db 77 PTSFQSETK 85

RESULT 15

ABB69700

ID ABB69700 standard; protein; 457 AA.

XX

AC ABB69700;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 35892.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL13803.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.

XX

PS Disclosure; SEQ ID NO 35892; 21pp + Sequence Listing; English.

XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 457 AA;

Query Match 72.1%; Score 31; DB 4; Length 457;  
Best Local Similarity 85.7%; Pred. No. 1e+03;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTSFNXA 7  
||| |  
Db 114 PTSFNGA 120

Search completed: February 10, 2005, 15:48:37  
Job time : 80.4648 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 20.1549 Seconds  
(without alignments)  
33.334 Million cell updates/sec

Title: US-10-067-484-2  
Perfect score: 43  
Sequence: 1 PTSFNXATK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	74.4	232	4	US-09-489-039A-7297	Sequence 7297, Ap
2	30	69.8	100	4	US-09-198-452A-502	Sequence 502, App
3	30	69.8	154	4	US-09-489-039A-12010	Sequence 12010, A
4	30	69.8	214	4	US-09-134-000C-4529	Sequence 4529, Ap
5	30	69.8	287	4	US-09-252-991A-23091	Sequence 23091, A
6	30	69.8	441	4	US-09-328-352-6369	Sequence 6369, Ap
7	30	69.8	987	4	US-09-540-236-3017	Sequence 3017, Ap
8	30	69.8	1380	4	US-09-328-352-8132	Sequence 8132, Ap
9	29	67.4	92	4	US-09-248-796A-17403	Sequence 17403, A
10	29	67.4	106	2	US-08-800-198-4	Sequence 4, Appli
11	29	67.4	106	3	US-09-296-595-4	Sequence 4, Appli
12	29	67.4	231	4	US-09-248-796A-15529	Sequence 15529, A
13	29	67.4	239	2	US-07-956-399-4	Sequence 4, Appli
14	29	67.4	240	2	US-08-800-198-8	Sequence 8, Appli
15	29	67.4	240	3	US-09-296-595-8	Sequence 8, Appli
16	29	67.4	292	4	US-09-328-352-6267	Sequence 6267, Ap
17	29	67.4	302	4	US-09-248-796A-14926	Sequence 14926, A
18	29	67.4	340	3	US-09-134-001C-5182	Sequence 5182, Ap
19	29	67.4	466	2	US-08-432-016-4	Sequence 4, Appli
20	29	67.4	466	2	US-08-684-594-4	Sequence 4, Appli
21	29	67.4	756	4	US-09-248-796A-19209	Sequence 19209, A
22	29	67.4	785	4	US-09-902-540-10007	Sequence 10007, A
23	29	67.4	935	4	US-09-134-000C-6493	Sequence 6493, Ap
24	29	67.4	1514	2	US-08-853-310-4	Sequence 4, Appli
25	28	65.1	34	3	US-09-100-600A-6	Sequence 6, Appli
26	28	65.1	34	3	US-09-100-600A-7	Sequence 7, Appli
27	28	65.1	34	3	US-09-100-600A-12	Sequence 12, Appl
28	28	65.1	34	3	US-09-100-600A-13	Sequence 13, Appl
29	28	65.1	34	3	US-09-100-600A-20	Sequence 20, Appl
30	28	65.1	34	3	US-09-100-600A-21	Sequence 21, Appl
31	28	65.1	36	3	US-09-100-600A-4	Sequence 4, Appli
32	28	65.1	36	3	US-09-100-600A-5	Sequence 5, Appli
33	28	65.1	36	3	US-09-100-600A-10	Sequence 10, Appl
34	28	65.1	36	3	US-09-100-600A-11	Sequence 11, Appl
35	28	65.1	45	3	US-09-100-600A-87	Sequence 87, Appl
36	28	65.1	45	3	US-09-100-600A-88	Sequence 88, Appl
37	28	65.1	45	3	US-09-100-600A-89	Sequence 89, Appl
38	28	65.1	45	3	US-09-100-600A-90	Sequence 90, Appl
39	28	65.1	45	3	US-09-100-600A-91	Sequence 91, Appl
40	28	65.1	46	3	US-09-100-600A-41	Sequence 41, Appl
41	28	65.1	46	4	US-09-270-767-37739	Sequence 37739, A
42	28	65.1	46	4	US-09-270-767-52956	Sequence 52956, A
43	28	65.1	47	3	US-09-100-600A-67	Sequence 67, Appl
44	28	65.1	47	3	US-09-100-600A-68	Sequence 68, Appl
45	28	65.1	47	3	US-09-100-600A-70	Sequence 70, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-489-039A-7297

; Sequence 7297, Application US/09489039A

; Patent No. 6610836

```
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7297
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7297
```

```
Query Match          74.4%; Score 32; DB 4; Length 232;
Best Local Similarity 66.7%; Pred. No. 32;
Matches      6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 PTSFNXATK 9
        | ||| ||:
Db      157 PRSFNAATE 165
```

#### RESULT 2

```
US-09-198-452A-502
; Sequence 502, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides,
fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the
diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 502
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-502
```

```
Query Match          69.8%; Score 30; DB 4; Length 100;
Best Local Similarity 55.6%; Pred. No. 34;
Matches      5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 PTSFNXATK 9
        ||||: |:
Db      47 PTSFSSCTR 55
```

#### RESULT 3



US-09-489-039A-12010  
 ; Sequence 12010, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 12010  
 ; LENGTH: 154  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-12010

Query Match 69.8%; Score 30; DB 4; Length 154;  
 Best Local Similarity 75.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXAT 8  
 ||| | ||  
 Db 139 PTSVNSAT 146

#### RESULT 4

US-09-134-000C-4529  
 ; Sequence 4529, Application US/09134000C  
 ; Patent No. 6617156  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 032796-032  
 ; CURRENT APPLICATION NUMBER: US/09/134,000C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/055,778  
 ; PRIOR FILING DATE: 1997-08-15  
 ; NUMBER OF SEQ ID NOS: 6812  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 4529  
 ; LENGTH: 214  
 ; TYPE: PRT  
 ; ORGANISM: Enterococcus faecalis  
 US-09-134-000C-4529

Query Match 69.8%; Score 30; DB 4; Length 214;  
 Best Local Similarity 55.6%; Pred. No. 79;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||: || |:  
 Db 56 PTAFNSQTQ 64

RESULT 5

US-09-252-991A-23091

; Sequence 23091, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 23091  
 ; LENGTH: 287  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-23091

Query Match 69.8%; Score 30; DB 4; Length 287;  
 Best Local Similarity 55.6%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||:|: ||  
 Db 31 PTAFSSTTK 39

RESULT 6

US-09-328-352-6369

; Sequence 6369, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6369  
 ; LENGTH: 441  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6369

Query Match 69.8%; Score 30; DB 4; Length 441;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy            3 SFNXATK 9  
             ||| |||  
Db            338 SFNSATK 344

RESULT 7

US-09-540-236-3017  
; Sequence 3017, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 3017  
; LENGTH: 987  
; TYPE: PRT  
; ORGANISM: M.catarrhalis  
US-09-540-236-3017

Query Match                    69.8%; Score 30; DB 4; Length 987;  
Best Local Similarity       66.7%; Pred. No. 4.3e+02;  
Matches       6; Conservative    0; Mismatches    3; Indels       0; Gaps       0;

Qy            1 PTSFNXATK 9  
             | ||| ||  
Db            964 PESFNVLTk 972

RESULT 8

US-09-328-352-8132  
; Sequence 8132, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 8132  
; LENGTH: 1380  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-8132

Query Match                    69.8%; Score 30; DB 4; Length 1380;  
Best Local Similarity       66.7%; Pred. No. 6.3e+02;  
Matches       6; Conservative    0; Mismatches    3; Indels       0; Gaps       0;

Qy            1 PTSFNXATK 9

| | | | |  
Db      1357 PESFNVLTk 1365

RESULT 9

US-09-248-796A-17403  
; Sequence 17403, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA  
ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17403  
; LENGTH: 92  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17403

Query Match                  67.4%; Score 29; DB 4; Length 92;  
Best Local Similarity      66.7%; Pred. No. 51;  
Matches      6; Conservative      1; Mismatches      2; Indels      0; Gaps      0;

Qy          1 PTSFNXATK 9  
             | | : | | |  
Db          29 PTTRNRATK 37

RESULT 10

US-08-800-198-4  
; Sequence 4, Application US/08800198  
; Patent No. 5942602  
; GENERAL INFORMATION:  
; APPLICANT: WELS, WINFRIED S.  
; APPLICANT: SCHMIDT, MATHIAS  
; APPLICANT: VAKALOPOULOU, EVANGELIA  
; APPLICANT: SCHNEIDER, DOUGLAS  
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.  
; STREET: 2200 CLARENDON BLVD. SUITE 1400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: US  
; ZIP: 22201  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```

;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/800,198
;   FILING DATE:  13-FEB-1997
;   CLASSIFICATION:  530
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  HAMLET-KING, DIANA
;   REGISTRATION NUMBER:  33,302
;   REFERENCE/DOCKET NUMBER:  SCH 1576
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  703-243-6333
;   TELEFAX:  703-243-6410
;   INFORMATION FOR SEQ ID NO:  4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  106 amino acids
;   TYPE:  amino acid
;   STRANDEDNESS:
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   HYPOTHETICAL:  NO
;   FRAGMENT TYPE:  internal
US-08-800-198-4

```

```

Query Match          67.4%;  Score 29;  DB 2;  Length 106;
Best Local Similarity  55.6%;  Pred. No. 59;
Matches      5;  Conservative    1;  Mismatches    3;  Indels      0;  Gaps      0;

```

```

Qy      1 PTSFNXATK 9
      ||:|  ||
Db      95 PTTFGAGTK 103

```

```

RESULT 11
US-09-296-595-4
; Sequence 4, Application US/09296595A
; Patent No. 6129915
; GENERAL INFORMATION:
; APPLICANT: WELS, WINFRIED S.
; APPLICANT: SCHMIDT, MATHIAS
; APPLICANT: VAKALOPOULOU, EVANGELIA
; APPLICANT: SCHNEIDER, DOUGLAS
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
; FILE REFERENCE: SCH-1576 D1
; CURRENT APPLICATION NUMBER: US/09/296,595A
; CURRENT FILING DATE: 1999-04-23
; EARLIER APPLICATION NUMBER: 08/800,198
; EARLIER FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Murine sp.
US-09-296-595-4

```

```

Query Match          67.4%;  Score 29;  DB 3;  Length 106;

```

Best Local Similarity 55.6%; Pred. No. 59;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||:| ||  
Db 95 PTTFGAGTK 103

RESULT 12

US-09-248-796A-15529  
; Sequence 15529, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA  
ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 15529  
; LENGTH: 231  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-15529

Query Match 67.4%; Score 29; DB 4; Length 231;  
Best Local Similarity 75.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TSFNXATK 9  
|||| ||  
Db 12 TSFNFQTK 19

RESULT 13

US-07-956-399-4  
; Sequence 4, Application US/07956399  
; Patent No. 5876717  
; GENERAL INFORMATION:  
; APPLICANT: SHIMAMURA, TOSHIRO  
; APPLICANT: TAKI, SHINSUKE  
; APPLICANT: HAMURO, JUNJI  
; TITLE OF INVENTION: POLYPEPTIDES CAPABLE OF BINDING TO HEAVY  
; TITLE OF INVENTION: CHAINS OF IL-2 RECEPTORS  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia

```

; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/956,399
; FILING DATE: 19921005
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5876717man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-586-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-956-399-4

```

```

Query Match          67.4%; Score 29; DB 2; Length 239;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches      5; Conservative      1; Mismatches      3; Indels      0; Gaps      0;

```

```

Qy      1 PTSFNXATK 9
      ||:|  ||
Db      96 PTTFGSGTK 104

```

# RESULT 14

US-08-800-198-8

```

; Sequence 8, Application US/08800198
; Patent No. 5942602
; GENERAL INFORMATION:
; APPLICANT: WELS, WINFRIED S.
; APPLICANT: SCHMIDT, MATHIAS
; APPLICANT: VAKALOPOULOU, EVANGELIA
; APPLICANT: SCHNEIDER, DOUGLAS
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
; STREET: 2200 CLARENDON BLVD. SUITE 1400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: US
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/800,198
;   FILING DATE:  13-FEB-1997
;   CLASSIFICATION:  530
;   ATTORNEY/AGENT INFORMATION:
;   NAME:  HAMLET-KING, DIANA
;   REGISTRATION NUMBER:  33,302
;   REFERENCE/DOCKET NUMBER:  SCH 1576
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  703-243-6333
;   TELEFAX:  703-243-6410
;   INFORMATION FOR SEQ ID NO:  8:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH:  240 amino acids
;   TYPE:  amino acid
;   STRANDEDNESS:
;   TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   HYPOTHETICAL:  NO
;   FRAGMENT TYPE:  internal
US-08-800-198-8

```

```

Query Match          67.4%;  Score 29;  DB 2;  Length 240;
Best Local Similarity  55.6%;  Pred. No. 1.5e+02;
Matches      5;  Conservative    1;  Mismatches    3;  Indels      0;  Gaps      0;

```

```

Qy      1 PTSFNXATK 9
      ||:|  ||
Db      229 PTTFGAGTK 237

```

# RESULT 15

```

US-09-296-595-8
; Sequence 8, Application US/09296595A
; Patent No. 6129915
; GENERAL INFORMATION:
; APPLICANT: WELS, WINFRIED S.
; APPLICANT: SCHMIDT, MATHIAS
; APPLICANT: VAKALOPOULOU, EVANGELIA
; APPLICANT: SCHNEIDER, DOUGLAS
; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
; FILE REFERENCE: SCH-1576 D1
; CURRENT APPLICATION NUMBER: US/09/296,595A
; CURRENT FILING DATE: 1999-04-23
; EARLIER APPLICATION NUMBER: 08/800,198
; EARLIER FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Murine sp.
US-09-296-595-8

```

```

Query Match          67.4%;  Score 29;  DB 3;  Length 240;

```



Best Local Similarity 55.6%; Pred. No. 1.5e+02;  
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||:| ||  
Db 229 PTTFGAGTK 237

Search completed: February 10, 2005, 16:02:06  
Job time : 20.1549 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:49:10 ; Search time 53.8732 Seconds  
(without alignments)  
54.586 Million cell updates/sec

Title: US-10-067-484-2  
Perfect score: 43  
Sequence: 1 PTSFNXATK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41	95.3	9	14	US-10-067-484-2	Sequence 2, Appli
2	41	95.3	9	14	US-10-067-620-2	Sequence 2, Appli
3	35	81.4	58	15	US-10-424-599-280809	Sequence 280809,
4	34	79.1	424	15	US-10-282-122A-69178	Sequence 69178, A
5	33	76.7	101	16	US-10-437-963-158168	Sequence 158168,
6	33	76.7	1179	16	US-10-437-963-167415	Sequence 167415,
7	32	74.4	148	16	US-10-767-701-48958	Sequence 48958, A
8	32	74.4	158	15	US-10-425-114-47711	Sequence 47711, A
9	32	74.4	179	16	US-10-437-963-192813	Sequence 192813,
10	32	74.4	436	16	US-10-437-963-159333	Sequence 159333,
11	32	74.4	713	14	US-10-238-075-263	Sequence 263, App
12	32	74.4	808	16	US-10-408-765A-722	Sequence 722, App
13	31	72.1	69	16	US-10-437-963-143687	Sequence 143687,
14	31	72.1	83	16	US-10-437-963-131259	Sequence 131259,
15	31	72.1	669	16	US-10-437-963-126181	Sequence 126181,
16	31	72.1	1361	15	US-10-282-122A-48282	Sequence 48282, A
17	30	69.8	52	15	US-10-424-599-267836	Sequence 267836,
18	30	69.8	73	15	US-10-424-599-262875	Sequence 262875,
19	30	69.8	81	16	US-10-437-963-188517	Sequence 188517,
20	30	69.8	94	15	US-10-424-599-262105	Sequence 262105,
21	30	69.8	100	15	US-10-289-762-502	Sequence 502, App
22	30	69.8	109	15	US-10-424-599-194644	Sequence 194644,
23	30	69.8	371	15	US-10-369-493-6753	Sequence 6753, Ap
24	30	69.8	407	15	US-10-282-122A-65777	Sequence 65777, A
25	30	69.8	418	15	US-10-362-881-31	Sequence 31, Appl
26	30	69.8	426	15	US-10-282-122A-65458	Sequence 65458, A
27	30	69.8	1362	15	US-10-282-122A-44993	Sequence 44993, A
28	30	69.8	1363	15	US-10-282-122A-63099	Sequence 63099, A
29	29	67.4	65	15	US-10-424-599-151501	Sequence 151501,
30	29	67.4	71	15	US-10-424-599-232397	Sequence 232397,
31	29	67.4	108	15	US-10-424-599-203533	Sequence 203533,
32	29	67.4	113	15	US-10-374-600-20	Sequence 20, Appl
33	29	67.4	113	15	US-10-374-531-20	Sequence 20, Appl
34	29	67.4	116	16	US-10-767-701-37364	Sequence 37364, A
35	29	67.4	127	15	US-10-374-600-5	Sequence 5, Appli
36	29	67.4	127	15	US-10-374-600-11	Sequence 11, Appl
37	29	67.4	127	15	US-10-374-600-15	Sequence 15, Appl
38	29	67.4	127	15	US-10-374-531-5	Sequence 5, Appli
39	29	67.4	127	15	US-10-374-531-11	Sequence 11, Appl
40	29	67.4	127	15	US-10-374-531-15	Sequence 15, Appl
41	29	67.4	233	15	US-10-282-122A-71928	Sequence 71928, A
42	29	67.4	296	15	US-10-425-114-51353	Sequence 51353, A
43	29	67.4	413	15	US-10-282-122A-50763	Sequence 50763, A
44	29	67.4	417	15	US-10-259-194A-212	Sequence 212, App

## ALIGNMENTS

## RESULT 1

US-10-067-484-2

; Sequence 2, Application US/10067484  
; Publication No. US20030170763A1  
; GENERAL INFORMATION:  
; APPLICANT: Buchanan, Bob B.  
; APPLICANT: del Val, Gregorio  
; APPLICANT: Frick, Oscar L.  
; TITLE OF INVENTION: RAGWEED ALLERGENS  
; FILE REFERENCE: 416272000200  
; CURRENT APPLICATION NUMBER: US/10/067,484  
; CURRENT FILING DATE: 2002-02-04  
; PRIOR APPLICATION NUMBER: US 60/266,686  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Ragweed  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 6  
; OTHER INFORMATION: Xaa= Leucine or Isoleucine  
US-10-067-484-2

Query Match 95.3%; Score 41; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||| |||||  
Db 1 PTSFNXATK 9

## RESULT 2

US-10-067-620-2

; Sequence 2, Application US/10067620  
; Publication No. US20030180225A1  
; GENERAL INFORMATION:  
; APPLICANT: Buchanan, Bob B.  
; APPLICANT: del Val, Gregorio  
; APPLICANT: Frick, Oscar L.  
; APPLICANT: Teuber, Suzanne S.  
; TITLE OF INVENTION: WALNUT AND RYEGRASS ALLERGENS  
; FILE REFERENCE: 416272003400  
; CURRENT APPLICATION NUMBER: US/10/067,620  
; CURRENT FILING DATE: 2002-02-04  
; PRIOR APPLICATION NUMBER: US 60/266,686  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Ragweed  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 6  
; OTHER INFORMATION: Xaa= Leucine or Isoleucine  
US-10-067-620-2

Query Match 95.3%; Score 41; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|||  
Db 1 PTSFNXATK 9

RESULT 3

US-10-424-599-280809  
; Sequence 280809, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 280809  
; LENGTH: 58  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(58)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_95593C.1.pep  
US-10-424-599-280809

Query Match 81.4%; Score 35; DB 15; Length 58;  
Best Local Similarity 66.7%; Pred. No. 4.6;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|::|||  
Db 32 PSNFNTATK 40

RESULT 4

US-10-282-122A-69178

; Sequence 69178, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 69178  
 ; LENGTH: 424  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 US-10-282-122A-69178

Query Match 79.1%; Score 34; DB 15; Length 424;  
 Best Local Similarity 66.7%; Pred. No. 63;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||||| |:  
 Db 368 PTSFNSVTE 376

RESULT 5

US-10-437-963-158168  
 ; Sequence 158168, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
 Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 158168  
 ; LENGTH: 101  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_57669C.1.pep  
 US-10-437-963-158168

Query Match 76.7%; Score 33; DB 16; Length 101;  
 Best Local Similarity 66.7%; Pred. No. 23;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 | ||| ||:  
 Db 42 PLSFNSATR 50

RESULT 6

US-10-437-963-167415  
 ; Sequence 167415, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
 Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 167415  
; LENGTH: 1179  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66029C.1.pep  
US-10-437-963-167415

Query Match 76.7%; Score 33; DB 16; Length 1179;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||||| : |  
Db 964 PTSFNSSKK 972

RESULT 7

US-10-767-701-48958  
; Sequence 48958, Application US/10767701  
; Publication No. US20040172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 48958  
; LENGTH: 148  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(148)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3476-023-P1-K1-D3.pep  
US-10-767-701-48958

Query Match 74.4%; Score 32; DB 16; Length 148;  
Best Local Similarity 77.8%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||| | |||  
Db 48 PTSKNVATK 56

RESULT 8

US-10-425-114-47711

; Sequence 47711, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated  
 With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 47711  
 ; LENGTH: 158  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 700051635\_FLI.pep  
 US-10-425-114-47711

Query Match 74.4%; Score 32; DB 15; Length 158;  
 Best Local Similarity 66.7%; Pred. No. 59;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 |:||| ||  
 Db 51 PSSFNKLTk 59

# RESULT 9

US-10-437-963-192813  
 ; Sequence 192813, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
 Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 192813  
 ; LENGTH: 179  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa



```
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(179)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_89008C.1.pep
US-10-437-963-192813
```

```
Query Match          74.4%; Score 32; DB 16; Length 179;
Best Local Similarity 85.7%; Pred. No. 68;
Matches      6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1 PTSFNXA 7
        |||||
Db      60 PTSFNSA 66
```

# RESULT 10

```
US-10-437-963-159333
; Sequence 159333, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 159333
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_5871C.1.pep
US-10-437-963-159333
```

```
Query Match          74.4%; Score 32; DB 16; Length 436;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches      6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 PTSFNXAT 8
        ||. || ||
Db     411 PTRFNAAT 418
```

# RESULT 11

```
US-10-238-075-263
; Sequence 263, Application US/10238075
```

```

; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which
are isolated from
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and
of their polypeptides.
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 263
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-263

```

```

Query Match          74.4%; Score 32; DB 14; Length 713;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches      6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      1 PTSFNXATK 9
        | ||| ||:
Db      140 PRSFNAATE 148

```

# RESULT 12

```

US-10-408-765A-722
; Sequence 722, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 722
; LENGTH: 808
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-722

```

```

Query Match          74.4%; Score 32; DB 16; Length 808;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches      6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 PTSFNXATK 9  
||:|| | |  
Db 489 PTNFNVAEK 497

RESULT 13

US-10-437-963-143687

; Sequence 143687, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 143687

; LENGTH: 69

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_44571C.1.pep

US-10-437-963-143687

Query Match 72.1%; Score 31; DB 16; Length 69;  
Best Local Similarity 75.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXAT 8  
||||| |  
Db 45 PTSFNHIT 52

RESULT 14

US-10-437-963-131259

; Sequence 131259, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 131259  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_33341C.1.pep  
US-10-437-963-131259

Query Match 72.1%; Score 31; DB 16; Length 83;  
Best Local Similarity 75.0%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXAT 8  
|| || ||  
Db 29 PTRFNIAT 36

# RESULT 15

US-10-437-963-126181  
; Sequence 126181, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 126181  
; LENGTH: 669  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_28753C.1.pep  
US-10-437-963-126181

Query Match 72.1%; Score 31; DB 16; Length 669;  
Best Local Similarity 85.7%; Pred. No. 4.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTSFNXA 7

|||||  
Db 598 PTSFNEA 604

Search completed: February 10, 2005, 16:41:30  
Job time : 54.8732 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 13.9437 Seconds  
(without alignments)  
62.104 Million cell updates/sec

Title: US-10-067-484-2  
Perfect score: 43  
Sequence: 1 PTSFNXATK 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	35	81.4	336	2	T09133	heat shock protein	
2	33	76.7	345	2	B43731	achaete-scute comp	
3	33	76.7	605	2	S18648	protein kinase wis	
4	33	76.7	779	1	WMVZAL	ribonucleoside-dip	
5	32	74.4	405	2	A75105	hypothetical prote	
6	32	74.4	713	2	E91118	probable ferrichro	
7	32	74.4	713	2	D85963	probable iron comp	

8	31	72.1	127	2	G75086	hypothetical prote
9	31	72.1	372	2	T39649	cell division cont
10	31	72.1	1300	2	T18364	ro-3 protein - Neu
11	31	72.1	1347	2	T41321	BTB domain and Ank
12	30	69.8	307	2	G69501	UDP-glucose 4-epim
13	30	69.8	362	2	AI0433	trypsin-like prote
14	30	69.8	371	2	T16391	hypothetical prote
15	30	69.8	399	2	AD2559	hypothetical prote
16	30	69.8	407	2	E81914	probable transmemb
17	30	69.8	414	2	C89428	protein T08D2.7 [i
18	30	69.8	426	2	F81187	glucose/galactose
19	30	69.8	498	2	H97214	endoglucanase, fam
20	30	69.8	529	2	AH0453	bifunctional purin
21	30	69.8	638	2	AE1483	B. subtilis Iold p
22	30	69.8	687	2	D84126	penicillin-binding
23	30	69.8	870	1	GNMVJA	pol polyprotein -
24	30	69.8	1029	2	H86179	hypothetical prote
25	29	67.4	178	2	S51388	hypothetical prote
26	29	67.4	202	2	A86864	conserved hypothet
27	29	67.4	213	2	A87259	hypothetical prote
28	29	67.4	216	2	AH2635	bacteriophage repr
29	29	67.4	216	2	G97417	hypothetical prote
30	29	67.4	256	2	C90443	hypothetical prote
31	29	67.4	374	2	S53829	ribosomal protein
32	29	67.4	389	2	S68175	cone arrestin - bu
33	29	67.4	389	2	S68172	cone arrestin - no
34	29	67.4	470	2	T15196	hypothetical prote
35	29	67.4	480	2	B64308	hypothetical prote
36	29	67.4	494	2	D64944	probable permease
37	29	67.4	494	2	F85794	probable transport
38	29	67.4	494	2	B90946	probable transport
39	29	67.4	498	1	HJBEI1	helicase (EC 3.6.1
40	29	67.4	523	2	I50478	neurolin - goldfis
41	29	67.4	527	2	D87318	conserved hypothet
42	29	67.4	743	2	T42557	tegument protein 1
43	29	67.4	903	2	JE0327	dynammin-related pr
44	29	67.4	903	2	T50334	dynammin-related pr
45	29	67.4	1350	2	T10803	probable RNA-direc

# ALIGNMENTS

## RESULT 1

T09133

heat shock protein homolog DNAJ - Trypanosoma brucei

N;Alternate names: chaperone

C;Species: Trypanosoma brucei

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004

C;Accession: T09133

R;Bringaud, F.; Vedrenne, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.; Venegas, J.; Merlin, G.; Baltz, T.

Mol. Biochem. Parasitol. 94, 249-264, 1998

A;Title: Conserved organization of genes in trypanosomatids.

A;Reference number: Z16580; MUID:98418771; PMID:9747975

A;Accession: T09133

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA  
A;Residues: 1-336 <BRI>  
A;Cross-references: UNIPROT:O76224; EMBL:AF031926; NID:g3452211;  
PIDN:AAC32771.1; PID:g3452212  
A;Experimental source: strain AnTat1  
C;Genetics:  
A;Gene: dnaJ  
C;Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
C;Keywords: heat shock; molecular chaperone; stress-induced protein  
F;4-70/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 81.4%; Score 35; DB 2; Length 336;  
Best Local Similarity 77.8%; Pred. No. 4.2;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||| | |||  
Db 320 PTSLNEATK 328

#### RESULT 2

B43731

achaete-scute complex protein T4 - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 03-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 09-Jul-2004

C;Accession: B43731; S35425

R;Villares, R.; Cabrera, C.V.

Cell 50, 415-424, 1987

A;Title: The achaete-scute gene complex of *Drosophila melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to myc.

A;Reference number: A43731; MUID:87273503; PMID:3111716

A;Accession: B43731

A;Molecule type: DNA

A;Residues: 1-345 <VIL1>

A;Cross-references: UNIPROT:P10084; GB:M17119

R;Villares, R.

submitted to the EMBL Data Library, November 1990

A;Reference number: S35425

A;Accession: S35425

A;Molecule type: DNA

A;Residues: 1-255, 'C', 257-345 <VIL2>

A;Cross-references: EMBL:M17119; NID:g156745; PID:g156748

C;Genetics:

A;Gene: FlyBase:sc

A;Cross-references: FlyBase:FBgn0004170

Query Match 76.7%; Score 33; DB 2; Length 345;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||: | |||  
Db 25 PTTINSATK 33

#### RESULT 3

S18648

protein kinase wis1 (EC 2.7.1.-) - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 22-Nov-1993 #sequence\_revision 10-Feb-1995 #text\_change 16-Aug-2004

C;Accession: S18648; T40435

R;Warbrick, E.; Fantes, P.A.

EMBO J. 10, 4291-4299, 1991

A;Title: The wis1 protein kinase is a dosage-dependent regulator of mitosis in *Schizosaccharomyces pombe*.

A;Reference number: S18648; MUID:92097549; PMID:1756736

A;Accession: S18648

A;Molecule type: DNA

A;Residues: 1-605 <WAR>

A;Cross-references: UNIPROT:P33886; EMBL:X62631; NID:g5141; PIDN:CAA44499.1; PID:g5142

R;Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Chillingworth, T.; Churcher, C.M. submitted to the EMBL Data Library, August 1999

A;Reference number: Z21929

A;Accession: T40435

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-605 <LYN>

A;Cross-references: EMBL:AL109822; PIDN:CAB52609.1; GSPDB:GN00067;

SPDB:SPBC409.07c

A;Experimental source: strain 972h-; cosmid c409

C;Genetics:

A;Gene: wis1

A;Map position: 2

C;Function:

A;Description: phosphotransferase

C;Superfamily: protein kinase homology

C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase

F;318-579/Domain: protein kinase homology <KIN>

F;326-334/Region: protein kinase ATP-binding motif

Query Match 76.7%; Score 33; DB 2; Length 605;

Best Local Similarity 66.7%; Pred. No. 23;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9

||||| |:

Db 224 PTSFNRQTR 232

#### RESULT 4

WMVZAL

ribonucleoside-diphosphate reductase (EC 1.17.4.1) large chain - African swine fever virus (strain Malawi LIL20/1)

N;Alternate names: ribonucleotide reductase large chain

C;Species: African swine fever virus, ASFV

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004

C;Accession: A40568

R;Boursnell, M.; Shaw, K.; Yanez, R.J.; Vinuela, E.; Dixon, L.

Virology 184, 411-416, 1991



A;Title: The sequences of the ribonucleotide reductase genes from African swine fever virus show considerable homology with those of the orthopoxvirus, vaccinia virus.

A;Reference number: A40568; MUID:91335775; PMID:1871976

A;Accession: A40568

A;Molecule type: DNA

A;Residues: 1-779 <BOU>

A;Cross-references: UNIPROT:P26685; GB:M64728; NID:g210649; PIDN:AAA42732.1; PID:g554615

C;Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain

C;Keywords: deoxyribonucleotide biosynthesis; early protein; oxidoreductase; redox-active disulfide

F;194-440,774-777/Disulfide bonds: redox-active #status predicted

F;420,424/Active site: Asn, Glu #status predicted

F;422/Active site: Cys (cysteine thiyl radical intermediate) #status predicted

Query Match 76.7%; Score 33; DB 1; Length 779;  
Best Local Similarity 66.7%; Pred. No. 30;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|| || ||  
Db 179 PTMFNAGTK 187

#### RESULT 5

A75105

hypothetical protein PAB1562 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004

C;Accession: A75105

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.

A;Reference number: A75001

A;Accession: A75105

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-405 <KAW>

A;Cross-references: UNIPROT:Q9UZH7; GB:AJ248286; GB:AL096836; NID:g5458366;

PIDN:CAB50142.1; PID:e1516039; PID:g5458654

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB1562

Query Match 74.4%; Score 32; DB 2; Length 405;  
Best Local Similarity 66.7%; Pred. No. 25;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|||||  
Db 265 PTSFNIAIK 273

#### RESULT 6

E91118

probable ferrichrome iron receptor precursor [imported] - Escherichia coli  
 (strain O157:H7, substrain RIMD 0509952)  
 C;Species: Escherichia coli  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C;Accession: E91118  
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.;  
 Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida,  
 T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara,  
 S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7  
 and genomic comparison with a laboratory strain K-12.  
 A;Reference number: A99629; MUID:21156231; PMID:11258796  
 A;Accession: E91118  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-713 <HAY>  
 A;Cross-references: UNIPROT:Q8XBQ5; GB:BA000007; PIDN:BAB37340.1; PID:g13363390;  
 GSPDB:GN00154  
 A;Experimental source: strain O157:H7, substrain RIMD 0509952  
 C;Genetics:  
 A;Gene: ECs3917

Query Match 74.4%; Score 32; DB 2; Length 713;  
 Best Local Similarity 66.7%; Pred. No. 46;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 | | | | |  
 Db 140 PRSFNAATE 148

#### RESULT 7

D85963

probable iron compound receptor Z4386 [imported] - Escherichia coli (strain  
 O157:H7, substrain EDL933)  
 C;Species: Escherichia coli  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C;Accession: D85963  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose,  
 D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.;  
 Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis,  
 N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, J.; Anantharaman, T.S.;  
 Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: D85963  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-713 <STO>  
 A;Cross-references: UNIPROT:Q8XBQ5; GB:AE005174; NID:g12517607; PIDN:AAG58168.1;  
 GSPDB:GN00145; UWGP:Z4386  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: Z4386

Query Match 74.4%; Score 32; DB 2; Length 713;  
Best Local Similarity 66.7%; Pred. No. 46;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
| ||| ||:  
Db 140 PRSFNAATE 148

RESULT 8

G75086

hypothetical protein PAB1650 - *Pyrococcus abyssi* (strain Orsay)

C;Species: *Pyrococcus abyssi*

C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004

C;Accession: G75086

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome structure and evolution.

A;Reference number: A75001

A;Accession: G75086

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-127 <KAW>

A;Cross-references: UNIPROT:Q9UZR0; GB:AJ248286; GB:AL096836; NID:g5458366;

PIDN:CAB49996.1; PID:g5458508

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB1650

C;Superfamily: *Pyrococcus horikoshii* hypothetical protein PH1129

Query Match 72.1%; Score 31; DB 2; Length 127;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
| |||: ||  
Db 31 PTSFSRGTK 39

RESULT 9

T39649

cell division control protein 27 - fission yeast (*Schizosaccharomyces pombe*)

C;Species: *Schizosaccharomyces pombe*

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C;Accession: T39649; T40271; S20487

R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.

submitted to the EMBL Data Library, October 1998

A;Reference number: Z21868

A;Accession: T39649

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-372 <LYN>

A;Cross-references: UNIPROT:P30261; EMBL:AL031856; PIDN:CAA21296.1;

GSPDB:GN00067; SPDB:SPBC1734.02c

A;Experimental source: strain 972h-; cosmid c1734

A;Accession: T40271

A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-372 <LY2>  
 A;Cross-references: EMBL:AL031854; PIDN:CAA21288.1; GSPDB:GN00067;  
 SPDB:SPBC337.18c  
 A;Experimental source: strain 972h-; cosmid c337  
 R;Hughes, D.A.; MacNeill, S.A.; Fantes, P.A.  
 Mol. Gen. Genet. 231, 401-410, 1992  
 A;Title: Molecular cloning and sequence analysis of cdc27(+) required for the  
 G(2)-M transition in the fission yeast *Schizosaccharomyces pombe*.  
 A;Reference number: S20487; MUID:92167959; PMID:1538696  
 A;Accession: S20487  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-17,19-372 <HUG>  
 C;Genetics:  
 A;Gene: SPBC1734.02c; SPBC337.18c  
 A;Map position: 2  
 A;Introns: 18/3; 37/2; 93/1; 115/3; 151/3

Query Match 72.1%; Score 31; DB 2; Length 372;  
 Best Local Similarity 66.7%; Pred. No. 38;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||: ||||  
 Db 343 PTTVNIATK 351

# RESULT 10

T18364

ro-3 protein - *Neurospora crassa*

C;Species: *Neurospora crassa*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T18364

R;Tinsley, J.H.; Minke, P.F.; Bruno, K.S.; Plamann, M.

submitted to the EMBL Data Library, November 1995

A;Description: Dynactin, a nonessential complex in *Neurospora*, is required for nuclear distribution.

A;Reference number: Z18895

A;Accession: T18364

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1300 <TIN>

A;Cross-references: UNIPROT:Q01397; EMBL:L48661; NID:g1050296; PID:g1050297;  
 PIDN:AAA80458.1

C;Genetics:

A;Gene: ro-3

A;Introns: 75/3

Query Match 72.1%; Score 31; DB 2; Length 1300;  
 Best Local Similarity 55.6%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||:|| |:  
 Db 71 PTTFNSPTR 79

RESULT 11

T41321

BTB domain and Ankaryin repeat containing protein - fission yeast  
(Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C;Accession: T41321

R;Gwilliam, R.; Barrell, B.G.; Rajandream, M.A.; Wedler, H.; Wambutt, R.  
submitted to the EMBL Data Library, September 1998

A;Reference number: Z21987

A;Accession: T41321

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1347 <GWI>

A;Cross-references: UNIPROT:O74881; EMBL:AL031603; PIDN:CAA20916.1;

GSPDB:GN00068; SPDB:SPCC330.11

A;Experimental source: strain 972h-; cosmid c330

C;Genetics:

A;Gene: SPDB:SPCC330.11

A;Map position: 3

Query Match	72.1%;	Score 31;	DB 2;	Length 1347;
Best Local Similarity	66.7%;	Pred. No. 1.5e+02;		
Matches	6;	Conservative	1;	Mismatches 2; Indels 0; Gaps 0;

Qy	1	PTSFXN	ATK	9
		:		
Db	1212	PTSWNLLTK		1220

RESULT 12

G69501

UDP-glucose 4-epimerase (galE-2) homolog - Archaeoglobus fulgidus

C;Species: Archaeoglobus fulgidus

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C;Accession: G69501

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: G69501

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-307 <KLE>

A;Cross-references: UNIPROT:O28263; GB:AE000963; GB:AE000782; NID:g2689286;  
PIDN:AAB89234.1; PID:g2648515; TIGR:AF2016  
C;Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase  
homology

Query Match 69.8%; Score 30; DB 2; Length 307;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||:| |||  
Db 124 PTTYGATK 132

#### RESULT 13

AI0433

trypsin-like proteinase degS (EC 3.4.21.-) - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C;Accession: AI0433

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.;  
Prentice, M.B.; Sebahia, M.; James, K.D.; Churcher, C.; Mungall, K.L.; Baker,  
S.; Basham, D.; Bentley, S.D.; Brooks, K.; Cerdeno-Tarraga, A.M.; Chillingworth,  
T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Feltwell, T.; Hamlin, N.;  
Holroyd, S.; Jagels, K.; Leather, S.; Karlyshev, A.V.; Moule, S.; Oyston,  
P.C.F.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
Whitehead, S.; Barrell, B.G.  
Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AI0433

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-362 <KUR>

A;Cross-references: UNIPROT:Q8ZB57; GB:AL590842; PIDN:CAC92797.1; PID:g15981490;  
GSPDB:GN00175

C;Genetics:

A;Gene: degS

C;Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain  
homology; trypsin homology

C;Keywords: hydrolase; serine proteinase

Query Match 69.8%; Score 30; DB 2; Length 362;  
Best Local Similarity 55.6%; Pred. No. 61;  
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|||:| |:  
Db 45 PTSYNQAVR 53

#### RESULT 14

T16391

hypothetical protein F47F2.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C;Accession: T16391

R;Bentley, D.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of *C. elegans* cosmid F47F2.  
A;Reference number: Z18506  
A;Accession: T16391  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-371 <BEN>  
A;Cross-references: UNIPROT:Q20541; EMBL:U40943; NID:g1072202; PID:g1072204;  
PIDN:AAA81716.1; CESP:F47F2.1  
C;Genetics:  
A;Gene: CESP:F47F2.1  
A;Introns: 39/3; 70/1; 126/2; 156/2; 182/3; 214/2; 286/2; 331/3  
C;Superfamily: kinase-related transforming protein; protein kinase homology  
F;61-317/Domain: protein kinase homology <KIN>

Query Match 69.8%; Score 30; DB 2; Length 371;  
Best Local Similarity 66.7%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
| | | | |  
Db 278 PRSFNLA AK 286

# RESULT 15

AD2559

hypothetical protein all8067 [imported] - *Nostoc* sp. (strain PCC 7120) plasmid  
pCC7120gamma

C;Species: *Nostoc* sp. PCC 7120

A;Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C;Accession: AD2559

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Iriguchi, M.; Ishikawa, A.; Kawashima, K.; Kimura, T.; Kishida, Y.; Kohara, M.;  
Matsumoto, M.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.;  
Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing  
Cyanobacterium *Anabaena* sp. strain PCC 7120.

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AD2559

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-399 <KUR>

A;Cross-references: UNIPROT:Q8YK50; GB:AP003603; PIDN:BAB77397.1; PID:g17134840;  
GSPDB:GN00182

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: all8067

A;Genome: plasmid

Query Match 69.8%; Score 30; DB 2; Length 399;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9

||| | :|

Search completed: February 10, 2005, 15:59:22  
Job time : 15.9437 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

```
Run on:      February 10, 2005, 15:38:08 ; Search time 65.662 Seconds
              (without alignments)
              70.188 Million cell updates/sec
```

```
Title:      US-10-067-484-2
Perfect score: 43
Sequence:   1 PTSFNXATK 9
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :      UniProt_03:*
           1:  uniprot_sprot:*
           2:  uniprot_trembl:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	%					Query	
No.	Score	Match	Length	DB	ID		Description
1	35	81.4	336	2	O76224		O76224 trypanosoma
2	35	81.4	492	2	Q81NH2		Q81nh2 bacillus an
3	34	79.1	398	2	Q8ZWX5		Q8zwx5 pyrobaculum
4	34	79.1	434	2	Q893V9		Q893v9 clostridium
5	33	76.7	242	2	Q8KDZ1		Q8kdz1 chlorobium
6	33	76.7	271	2	Q94LA9		Q94la9 arabidopsis
7	33	76.7	336	2	O77029		O77029 drosophila
8	33	76.7	345	1	AST4_DROME		P10084 drosophila
9	33	76.7	346	2	O77031		O77031 drosophila



10	33	76.7	460	2	Q6C3D8	Q6c3d8	yarrowia li
11	33	76.7	468	2	Q758Z1	Q758z1	ashbya goss
12	33	76.7	605	1	WIS1_SCHPO	P33886	schizosacch
13	33	76.7	778	1	RIR1_ASFB7	P42491	african swi
14	33	76.7	779	1	RIR1_ASFM2	P26685	african swi
15	33	76.7	901	2	Q6CAJ2	Q6caj2	yarrowia li
16	33	76.7	1041	2	Q6X5T7	Q6x5t7	streptococc
17	32	74.4	185	2	Q9AY89	Q9ay89	oryza sativ
18	32	74.4	320	2	Q7RGP9	Q7rgp9	plasmodium
19	32	74.4	405	2	Q9UZB7	Q9uzb7	pyrococcus
20	32	74.4	459	2	Q64U70	Q64u70	bacteroides
21	32	74.4	492	2	Q639A8	Q639a8	bacillus ce
22	32	74.4	492	2	Q735B2	Q735b2	bacillus ce
23	32	74.4	492	2	Q6HGM1	Q6hgm1	bacillus th
24	32	74.4	713	2	Q8FDI8	Q8fdi8	escherichia
25	32	74.4	713	2	Q8XBQ5	Q8xbq5	escherichia
26	32	74.4	808	2	Q9UK88	Q9uk88	homo sapien
27	32	74.4	844	2	Q6P517	Q6p517	homo sapien
28	32	74.4	897	2	Q6FL60	Q6fl60	candida gla
29	32	74.4	1434	2	Q8IJI3	Q8iji3	plasmodium
30	32	74.4	1940	2	Q7SAX4	Q7sax4	neurospora
31	32	74.4	2829	2	P70039	P70039	xenopus lae
32	32	74.4	3347	2	Q8MMJ9	Q8mmj9	bombyx mori
33	32	74.4	3354	2	Q8T101	Q8t101	bombyx mori
34	31	72.1	88	2	Q8JUI4	Q8jui4	foot-and-mo
35	31	72.1	124	2	Q949J0	Q949j0	cucumis sat
36	31	72.1	127	2	Q9UZR0	Q9uzr0	pyrococcus
37	31	72.1	133	2	Q6PUC2	Q6puc2	anopheles g
38	31	72.1	135	2	Q88CV7	Q88cv7	pseudomonas
39	31	72.1	156	2	Q6F173	Q6f173	mesoplasma
40	31	72.1	211	2	Q952X5	Q952x5	partula toh
41	31	72.1	211	2	Q952X8	Q952x8	partula tae
42	31	72.1	211	2	Q952Y3	Q952y3	partula moo
43	31	72.1	211	2	Q952Y4	Q952y4	partula moo
44	31	72.1	221	2	Q872B6	Q872b6	neurospora
45	31	72.1	262	2	Q8DLX3	Q8dlx3	synechococc

#### ALIGNMENTS

#### RESULT 1

O76224

ID O76224 PRELIMINARY; PRT; 336 AA.  
AC O76224;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Chaperone.  
GN Name=DNAJ;  
OS Trypanosoma brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5691;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AnTat1;  
RX MEDLINE=98418771; PubMed=9747975; DOI=10.1016/S0166-6851(98)00080-2;

RA Bringaud F., Vedrenne C., Cuvillier A., Parzy D., Baltz D., Tetaud E.,  
 RA Pays E., Venegas J., Merlin G., Baltz T.;  
 RT "Conserved organization of genes in trypanosomatids.";  
 RL Mol. Biochem. Parasitol. 94:249-264(1998).  
 DR EMBL; AF031926; AAC32771.1; -.  
 DR PIR; T09133; T09133.  
 DR HSSP; P25685; 1HDJ.  
 DR GO; GO:0051082; F:unfolded protein binding; IEA.  
 DR GO; GO:0006457; P:protein folding; IEA.  
 DR InterPro; IPR002939; DnaJ\_C.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR008971; HSP40\_DnaJ\_pep.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ; 1.  
 DR Pfam; PF01556; DnaJ\_C; 1.  
 DR PRINTS; PR00625; DNAJPROTEIN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR PROSITE; PS00636; DNAJ\_1; 1.  
 DR PROSITE; PS50076; DNAJ\_2; 1.  
 KW Chaperone.  
 SQ SEQUENCE 336 AA; 36435 MW; 18BD9332E3B0F0EF CRC64;

Query Match 81.4%; Score 35; DB 2; Length 336;  
 Best Local Similarity 77.8%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9  
 ||| | |||  
 Db 320 PTSLNEATK 328

## RESULT 2

Q81NH2

ID Q81NH2 PRELIMINARY; PRT; 492 AA.  
 AC Q81NH2; Q6HWN7; Q6KQS9;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Drug resistance transporter, EmrB/QacA family.  
 GN OrderedLocusNames=BA3223, BAS2994, GBAA3223;  
 OS Bacillus anthracis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
 RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
 RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,  
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,  
 RA Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M.,

RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria."  
 RL Nature 423:81-86(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
 RA Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics."  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sterne;  
 RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AE017034; AAP27016.1; -.  
 DR EMBL; AE017334; AAT32338.1; -.  
 DR EMBL; AE017225; AAT55302.1; -.  
 DR TIGR; BA3223; -.  
 DR TIGR; GBAA3223; -.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0015520; F:tetracycline:hydrogen antiporter activity; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0015904; P:tetracycline transport; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR004638; Efflux\_EmrB.  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR001411; TCR\_TetB.  
 DR PRINTS; PR01036; TCRTETB.  
 DR TIGRFAMs; TIGR00711; efflux\_EmrB; 1.  
 DR PROSITE; PS50850; MFS; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 492 AA; 53552 MW; C7D3B7C6487275CA CRC64;

Query Match 81.4%; Score 35; DB 2; Length 492;  
 Best Local Similarity 77.8%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 |||| |||  
 Db 436 PTSFTEATK 444

# RESULT 3

Q8ZWX5

ID Q8ZWX5 PRELIMINARY; PRT; 398 AA.  
 AC Q8ZWX5;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE PaREP2b.  
 GN OrderedLocusNames=PAE1574;  
 OS Pyrobaculum aerophilum.

OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869; DOI=10.1073/pnas.241636498;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL; AE009828; AAL63574.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 398 AA; 45218 MW; 3DC686B0A50123CE CRC64;  
  
 Query Match 79.1%; Score 34; DB 2; Length 398;  
 Best Local Similarity 66.7%; Pred. No. 44;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||:|| ||  
 Db 200 PTAFNAAVK 208

#### RESULT 4

Q893V9

ID Q893V9 PRELIMINARY; PRT; 434 AA.  
 AC Q893V9;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Putative phosphoenolpyruvate phosphomutase.  
 GN OrderedLocusNames=CTC01698;  
 OS Clostridium tetani.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Massachusetts / E88;  
 RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;  
 RA Brueggemann H., Baeumer S., Fricke W.F., Wiezer A., Liesegang H.,  
 RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,  
 RA Gottschalk G.;  
 RT "The genome sequence of Clostridium tetani, the causative agent of  
 RT tetanus disease.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).  
 DR EMBL; AE015942; AAO36233.1; -.  
 DR HSSP; P56839; 1PYM.  
 DR GO; GO:0016779; F:nucleotidyltransferase activity; IEA.  
 DR GO; GO:0009058; P:biosynthesis; IEA.  
 DR InterPro; IPR004820; Cytidylyltransf.  
 DR InterPro; IPR004821; Cyt\_trans\_rel.  
 DR Pfam; PF01467; CTP\_transf\_2; 1.  
 DR TIGRFAMs; TIGR00125; cyt\_tran\_rel; 1.  
 KW Complete proteome.

SQ SEQUENCE 434 AA; 49334 MW; 9F5EC8A0C82FA3BA CRC64;

Query Match 79.1%; Score 34; DB 2; Length 434;  
Best Local Similarity 66.7%; Pred. No. 49;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
||||| |:  
Db 370 PTSFNTVTE 378

RESULT 5

Q8KDZ1

ID Q8KDZ1 PRELIMINARY; PRT; 242 AA.  
AC Q8KDZ1;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Transcriptional regulator, putative.  
GN OrderedLocusNames=CT0903;  
OS Chlorobium tepidum.  
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;  
OC Chlorobaculum.  
OX NCBI\_TaxID=1097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TLS / ATCC 49652 / DSM 12025;  
RX MEDLINE=22103685; PubMed=12093901; DOI=10.1073/pnas.132181499;  
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,  
RA Dodson R.J., DeBoy R.T., Gwinn M.L., Nelson W.C., Haft D.H.,  
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.F., Yang F.,  
RA Holt I.E., Umayam L.A., Mason T.M., Brenner M., Shea T.P.,  
RA Parksey D.S., Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B.,  
RA Radune D., Vamathevan J.J., Khouri H.M., White O., Gruber T.M.,  
RA Ketchum K.A., Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;  
RT "The complete genome sequence of Chlorobium tepidum TLS, a  
RT photosynthetic, anaerobic, green-sulfur bacterium."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).  
DR EMBL; AE012856; AAM72138.1; -.  
DR TIGR; CT0903; -.  
KW Complete proteome.  
SQ SEQUENCE 242 AA; 27147 MW; DB45122A9065D10C CRC64;

Query Match 76.7%; Score 33; DB 2; Length 242;  
Best Local Similarity 66.7%; Pred. No. 43;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|: || |||  
Db 27 PSKFNLATK 35

RESULT 6

Q94LA9

ID Q94LA9 PRELIMINARY; PRT; 271 AA.  
AC Q94LA9;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Hypothetical protein T18F15.4 (At1g44542).  
 GN Name=T18F15.4;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,  
 RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Carninci P., Hayashizaki Y.,  
 RA Ishida J., Kamiya A., Kawai J., Narusaka M., Sakurai T., Satou M.,  
 RA Seki M., Shinozaki K., Ecker J.R.;  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AC084807; AAK43483.1; -.  
 DR EMBL; BT012646; AAT06465.1; -.  
 DR InterPro; IPR007325; Cyclase.  
 DR Pfam; PF04199; Cyclase; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 271 AA; 30636 MW; 871911F62A9AB110 CRC64;  
  
 Query Match 76.7%; Score 33; DB 2; Length 271;  
 Best Local Similarity 66.7%; Pred. No. 49;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9  
 ||:|: |||  
 Db 30 PTTFSVATK 38

# RESULT 7

O77029  
 ID O77029 PRELIMINARY; PRT; 336 AA.  
 AC O77029;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Achaete-scute complex protein SC (Scute protein).  
 GN Name=sc;  
 OS Drosophila yakuba (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IVORY COAST;  
 RX MEDLINE=98278813; PubMed=9611206;  
 RA Takano T.S.;  
 RT "Rate variation of DNA sequence evolution in the Drosophila

RT lineages.";  
 RL Genetics 149:959-970(1998).  
 CC -!- FUNCTION: AS-C PROTEINS ARE INVOLVED IN THE DETERMINATION OF THE  
 CC NEURONAL PRECURSORS IN THE PERIPHERAL NERVOUS SYSTEM AND THE  
 CC CENTRAL NERVOUS SYSTEM. ALSO INVOLVED IN SEX DETERMINATION AND  
 CC DOSAGE COMPENSATION (BY SIMILARITY).  
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 DR EMBL; AB005799; BAA33210.1; -.  
 DR FlyBase; FBgn0025397; Dyak\sc.  
 DR GO; GO:0030154; P:cell differentiation; IEA.  
 DR GO; GO:0007399; P:neurogenesis; IEA.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS50888; HLH; 1.  
 KW Developmental protein; Differentiation; Neurogenesis.  
 FT DNA\_BIND 90 100 BASIC DOMAIN (BY SIMILARITY).  
 FT DOMAIN 101 151 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 336 AA; 37050 MW; 0202BB37BCB1A9BC CRC64;

Query Match 76.7%; Score 33; DB 2; Length 336;  
 Best Local Similarity 66.7%; Pred. No. 62;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9  
 ||: |||  
 Db 13 PTTINSATK 21

# RESULT 8

## AST4\_DROME

ID AST4\_DROME STANDARD; PRT; 345 AA.  
 AC P10084; O76890;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 25-JAN-2005 (Rel. 46, Last annotation update)  
 DE Achaete-scute complex protein T4 (Scute protein).  
 GN Name=sc; Synonyms=T4; ORFNames=CG3827;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Canton-S;  
 RX MEDLINE=87273503; PubMed=3111716; DOI=10.1016/0092-8674(87)90495-8;  
 RA Villares R., Cabrera C.V.;  
 RT "The achaete-scute gene complex of D. melanogaster: conserved domains  
 RT in a subset of genes required for neurogenesis and their homology to  
 RT myc.";  
 RL Cell 50:415-424(1987).  
 RN [2]  
 RP REVISIONS.

RA Villares R.;  
 RL Submitted (SEP-1988) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP GENOME REANNOTATION.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).



RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R;  
 RX MEDLINE=20196011; PubMed=10731137; DOI=10.1126/science.287.5461.2220;  
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,  
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,  
 RA Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of *D.  
 RT melanogaster*.";  
 RL Science 287:2220-2222 (2000).  
 RN [6]  
 RP FUNCTION.  
 RX MEDLINE=90059894; PubMed=2583094;  
 RA Torres M., Sanchez L.;  
 RT "The scute (T4) gene acts as a numerator element of the X: a signal  
 RT that determines the state of activity of sex-lethal in *Drosophila*.";  
 RL EMBO J. 8:3079-3086 (1989).  
 CC -!- FUNCTION: AS-C proteins are involved in the determination of the  
 CC neuronal precursors in the peripheral nervous system and the  
 CC central nervous system. Also involved in sex determination and  
 CC dosage compensation.  
 CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another  
 CC bHLH protein.  
 CC -!- TISSUE SPECIFICITY: L(1)SC, SC and AC strongly label the  
 CC presumptive stomatogastric nervous system, while ASE is more  
 CC prominent in the presumptive procephalic lobe.  
 CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M17119; AAA28313.1; -.  
 DR EMBL; AE003417; AAF45499.1; -..  
 DR EMBL; AL024453; CAA19657.1; -..  
 DR PIR; B43731; B43731.  
 DR IntAct; P10084; -..  
 DR TRANSFAC; T00004; -..  
 DR FlyBase; FBgn0004170; sc.  
 DR GO; GO:0008407; P:bristle morphogenesis; NAS.  
 DR GO; GO:0007399; P:neurogenesis; IGI.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
 DR GO; GO:0007530; P:sex determination; IMP.  
 DR GO; GO:0007540; P:sex determination, establishment of X:A ratio; NAS.  
 DR GO; GO:0007419; P:ventral cord development; NAS.  
 DR InterPro; IPR001092; HLH\_basic.

DR Pfam; PF00010; HLH; 1.  
 DR PROSITE; PS50888; HLH; 1.  
 KW Developmental protein; Differentiation; Neurogenesis.  
 FT DNA\_BIND 102 112 Basic motif.  
 FT DOMAIN 113 163 Helix-loop-helix motif.  
 FT CONFLICT 161 161 R -> S (in Ref. 1).  
 FT CONFLICT 213 213 T -> R (in Ref. 1).  
 FT CONFLICT 219 219 L -> V (in Ref. 1).  
 SQ SEQUENCE 345 AA; 38155 MW; DE68E49A8CCF16EB CRC64;

Query Match 76.7%; Score 33; DB 1; Length 345;  
 Best Local Similarity 66.7%; Pred. No. 64;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||: |||  
 Db 25 PTTINSATK 33

# RESULT 9

077031

ID 077031 PRELIMINARY; PRT; 346 AA.  
 AC 077031;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Achaete-scute complex protein SC (Scute protein).  
 GN Name=sc;  
 OS Drosophila simulans (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7240;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SIM-5 G20;  
 RX MEDLINE=98278813; PubMed=9611206;  
 RA Takano T.S.;  
 RT "Rate variation of DNA sequence evolution in the Drosophila  
 RT lineages."  
 RL Genetics 149:959-970 (1998).  
 CC -!- FUNCTION: AS-C PROTEINS ARE INVOLVED IN THE DETERMINATION OF THE  
 CC NEURONAL PRECURSORS IN THE PERIPHERAL NERVOUS SYSTEM AND THE  
 CC CENTRAL NERVOUS SYSTEM. ALSO INVOLVED IN SEX DETERMINATION AND  
 CC DOSAGE COMPENSATION (BY SIMILARITY).  
 CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 DR EMBL; AB005801; BAA33212.1; -.  
 DR FlyBase; FBgn0012893; Dsim\sc.  
 DR GO; GO:0030154; P:cell differentiation; IEA.  
 DR GO; GO:0007399; P:neurogenesis; IEA.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS50888; HLH; 1.

KW Developmental protein; Differentiation; Neurogenesis.  
 FT DNA\_BIND 102 112 BASIC DOMAIN (BY SIMILARITY).  
 FT DOMAIN 113 163 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 SQ SEQUENCE 346 AA; 38321 MW; 1433B75DBC0A534A CRC64;

Query Match 76.7%; Score 33; DB 2; Length 346;  
 Best Local Similarity 66.7%; Pred. No. 64;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PTSFNXATK 9  
 ||: | |||  
 Db 25 PTTINSATK 33

# RESULT 10

Q6C3D8

ID Q6C3D8 PRELIMINARY; PRT; 460 AA.  
 AC Q6C3D8;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Similarity.  
 GN ORFNames=YALIOF00550g;  
 OS Yarrowia lipolytica CLIB99.  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 OX NCBI\_TaxID=284591;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIB99;  
 RG Genolevures;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,  
 RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.L.;  
 RT "Genome evolution in yeasts."  
 RL Nature 430:35-44(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIB99;  
 RA Genoscope;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; CR382132; CAG77626.1; -.  
 SQ SEQUENCE 460 AA; 47253 MW; 6B589CBD93AA96C6 CRC64;

Query Match 76.7%; Score 33; DB 2; Length 460;  
 Best Local Similarity 75.0%; Pred. No. 88;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTSFNXAT 8  
|:| | | |  
Db 4 PSSFNTAT 11

RESULT 11

Q758Z1

ID Q758Z1 PRELIMINARY; PRT; 468 AA.  
AC Q758Z1;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE ADR386Wp.  
GN ORFNames=ADR386W;  
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.  
OX NCBI\_TaxID=33169;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 10895;  
RA Voegeli S.E., Brachat S., Dietrich F.S., Lerch A., Gaffney T.,  
RA Philippsen P.;  
RL Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE016817; AAS52306.1; -.  
DR ACD; ADR386W; -.  
SQ SEQUENCE 468 AA; 50321 MW; B790FFAAF472DE54 CRC64;

Query Match 76.7%; Score 33; DB 2; Length 468;  
Best Local Similarity 75.0%; Pred. No. 90;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PTSFNXAT 8  
|:| | | |  
Db 401 PSSFNAAT 408

RESULT 12

WIS1\_SCHPO

ID WIS1\_SCHPO STANDARD; PRT; 605 AA.  
AC P33886;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Protein kinase wis1 (EC 2.7.1.-) (Protein kinase sty2).  
GN Name=wis1; Synonyms=spc2, sty2; ORFNames=SPBC409.07c;  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=92097549; PubMed=1756736;

RA Warbrick E., Fantes P.A.;  
 RT "The wis1 protein kinase is a dosage-dependent regulator of mitosis in  
 RT Schizosaccharomyces pombe.";  
 RL EMBO J. 10:4291-4299(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: Dosage-dependent regulator of mitosis with serine/  
 CC threonine protein kinase activity. May play a role in the  
 CC integration of nutritional sensing with the control over entry  
 CC into mitosis. It may interact with cdc25, wee1 and win1. May  
 CC activate styl.  
 CC -!- PTM: Dephosphorylated by pyp1 and pyp2.  
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family. MAP  
 CC kinase kinase subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X62631; CAA44499.1; -.  
 DR EMBL; AL109822; CAB52609.1; -.  
 DR PIR; S18648; S18648.  
 DR HSSP; P35968; 1VR2.  
 DR GeneDB\_SPombe; SPBC409.07c; -.  
 DR InterPro; IPR011009; Kinase\_like.

DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Cell cycle; Cell division; Mitosis; Phosphorylation;  
 KW Serine/threonine-protein kinase; Transferase.  
 FT DOMAIN 320 579 Protein kinase.  
 FT NP\_BIND 326 334 ATP (By similarity).  
 FT BINDING 349 349 ATP (By similarity).  
 FT ACT\_SITE 441 441 Proton acceptor (By similarity).  
 FT MOD\_RES 469 469 Phosphoserine (By similarity).  
 FT MOD\_RES 473 473 Phosphothreonine (By similarity).  
 SQ SEQUENCE 605 AA; 64762 MW; 3EB97AF74190AD93 CRC64;

Query Match 76.7%; Score 33; DB 1; Length 605;  
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
 ||||| |:  
 Db 224 PTSFNRQTR 232

# RESULT 13

## RIR1\_ASFB7

ID RIR1\_ASFB7 STANDARD; PRT; 778 AA.  
 AC P42491;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Ribonucleoside-diphosphate reductase large chain (EC 1.17.4.1)  
 DE (Ribonucleotide reductase).  
 GN Name=F778R;  
 OS African swine fever virus (strain BA71V) (ASFV).  
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.  
 OX NCBI\_TaxID=10498;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=11831707;  
 RA Yanez R.J., Rodriguez J.M., Nogal M.I., Yuste L., Enriquez C.,  
 RA Rodriguez J.F., Vinuela E.;  
 RT "Analysis of the complete nucleotide sequence of African swine fever  
 RT virus."  
 RL Virology 208:249-278(1995).  
 CC -!- FUNCTION: Provides the precursors necessary for DNA synthesis.  
 CC -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate +  
 CC thioredoxin disulfide + H(2)O = ribonucleoside diphosphate +  
 CC thioredoxin.  
 CC -!- PATHWAY: DNA replication pathway; first step.  
 CC -!- SUBUNIT: Heterodimer of a large and a small chain.  
 CC -!- SIMILARITY: Belongs to the ribonucleoside diphosphate reductase  
 CC large chain family.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18466; AAA65275.1; -.
DR InterPro; IPR000788; Ribonucleo_red.
DR InterPro; IPR008926; Ribonucleo_red_N.
DR Pfam; PF02867; Ribonuc_red_lgC; 1.
DR Pfam; PF00317; Ribonuc_red_lgN; 1.
DR PRINTS; PR01183; RIBORDTASEM1.
DR PROSITE; PS00089; RIBORED_LARGE; 1.
KW DNA replication; Early protein; Oxidoreductase.
SQ SEQUENCE 778 AA; 87492 MW; 9DB88008677A877F CRC64;

```

```

Query Match          76.7%; Score 33; DB 1; Length 778;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches      6; Conservative      0; Mismatches      3; Indels      0; Gaps      0;

```

```

Qy      1 PTSFNXATK 9
        || || ||
Db      178 PTMFNAGTK 186

```

#### RESULT 14

##### RIR1\_ASF2

```

ID RIR1_ASF2 STANDARD; PRT; 779 AA.
AC P26635;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Ribonucleoside-diphosphate reductase large chain (EC 1.17.4.1)
DE (Ribonucleotide reductase).
OS African swine fever virus (isolate Malawi Lil 20/1) (ASFV).
OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
OX NCBI_TaxID=10500;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335775; PubMed=1871976;
RA Boursnell M., Shaw K., Yanez R.J., Vinuela E., Dixon L.;
RT "The sequences of the ribonucleotide reductase genes from African
RT swine fever virus show considerable homology with those of the
RT orthopoxvirus, vaccinia virus.";
RL Virology 184:411-416(1991).
CC -!- FUNCTION: Provides the precursors necessary for DNA synthesis.
CC -!- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate +
CC thioredoxin disulfide + H(2)O = ribonucleoside diphosphate +
CC thioredoxin.
CC -!- PATHWAY: DNA replication pathway; first step.
CC -!- SUBUNIT: Heterodimer of a large and a small chain.
CC -!- SIMILARITY: Belongs to the ribonucleoside diphosphate reductase
CC large chain family.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M64728; AAA42732.1; -.  
DR PIR; A40568; WMVZAL.  
DR InterPro; IPR000788; Ribonucleo\_red.  
DR InterPro; IPR008926; Ribonucleo\_red\_N.  
DR Pfam; PF02867; Ribonuc\_red\_lgC; 1.  
DR Pfam; PF00317; Ribonuc\_red\_lgN; 1.  
DR PRINTS; PR01183; RIBORDTASEM1.  
DR PROSITE; PS00089; RIBORED\_LARGE; 1.  
KW DNA replication; Early protein; Oxidoreductase.  
SQ SEQUENCE 779 AA; 87388 MW; 88A3D0C8D5D10819 CRC64;

Query Match 76.7%; Score 33; DB 1; Length 779;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PTSFNXATK 9  
|| || ||  
Db 179 PTMFNAGTK 187

#### RESULT 15

Q6CAJ2

ID Q6CAJ2 PRELIMINARY; PRT; 901 AA.  
AC Q6CAJ2;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Similar to sp|P08640 *Saccharomyces cerevisiae* YIR019c STA1  
DE extracellular alpha-1.  
GN ORFNames=YALI0D02299g;  
OS *Yarrowia lipolytica* CLIB99.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodascaceae; *Yarrowia*.  
OX NCBI\_TaxID=284591;  
RN [1]  
RF SEQUENCE FROM N.A.  
RC STRAIN=CLIB99;  
RG Genolevures;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,  
RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,  
RA Swennene D., Tekaia F., Wesolowski-Louvel M., Westhof E., Wirth B.,



RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
RA Wincker P., Souciet J.L.;  
RT "Genome evolution in yeasts.";  
RL Nature 430:35-44 (2004).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CLIB99;  
RA Genoscope;  
RL Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; CR382130; CAG80506.1; -.  
SQ SEQUENCE 901 AA; 90594 MW; 01E104DE8AB7EFF7 CRC64;

Query Match 76.7%; Score 33; DB 2; Length 901;  
Best Local Similarity 75.0%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PTSFNXAT 8  
|||||  
Db 878 PTSFNATT 885

Search completed: February 10, 2005, 15:57:22  
Job time : 67.662 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 78.4648 Seconds  
(without alignments)  
44.362 Million cell updates/sec

Title: US-10-067-484-3  
Perfect score: 44  
Sequence: 1 XYGLVQFNR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	43	97.7	9	5	ABB81970	Abb81970 30 kDa ra
2	33	75.0	99	4	AAU56543	Aau56543 Propionib
3	33	75.0	99	6	ABM53062	Abm53062 Propionib
4	33	75.0	170	6	ABU28434	Abu28434 Protein e
5	33	75.0	192	7	ADM26790	Adm26790 Hyperther
6	33	75.0	253	5	ABP25912	Abp25912 Streptoco
7	33	75.0	329	3	AAB51794	Aab51794 Gene 21 h
8	33	75.0	819	6	ADA33363	Ada33363 Acinetoba
9	32	72.7	112	4	ABB16566	Abb16566 Human ner

10	32	72.7	124	2	AAAY36180	Aay36180	Human	sec
11	32	72.7	124	2	AAAY36133	Aay36133	Human	sec
12	32	72.7	124	2	AAAY36210	Aay36210	Human	sec
13	32	72.7	124	4	AAM39501	Aam39501	Human	pol
14	32	72.7	124	5	ABB89828	Abb89828	Human	pol
15	32	72.7	124	5	AAM52201	Aam52201	Human	MP-
16	32	72.7	124	5	AAM52202	Aam52202	Mouse	MP-
17	32	72.7	124	7	ADJ46062	Adj46062	Novel	hum
18	32	72.7	124	7	ADJ45938	Adj45938	Novel	hum
19	32	72.7	124	7	ADJ46032	Adj46032	Novel	hum
20	32	72.7	124	8	ADP19441	Adp19441	Human	sec
21	32	72.7	124	8	ADP19471	Adp19471	Human	sec
22	32	72.7	132	2	AAAY76521	Aay76521	Human	ova
23	32	72.7	132	4	AAM41287	Aam41287	Human	pol
24	32	72.7	291	5	ABG91467	Abg91467	Purine/py	
25	32	72.7	294	8	ADS30672	Ads30672	Bacterial	
26	32	72.7	296	3	AAB15980	Aab15980	E. coli	p
27	32	72.7	296	6	ABU14692	Abu14692	Protein	e
28	32	72.7	323	2	AAW55360	Aaw55360	H. pylori	
29	32	72.7	323	2	AAAY17165	Aay17165	H. pylori	
30	32	72.7	476	2	AAW69946	Aaw69946	GlutRNAGl	
31	32	72.7	949	7	ADM25498	Adm25498	Hyperther	
32	32	72.7	1282	8	ADN46345	Adn46345	Thermococ	
33	31	70.5	203	4	AAU21601	Aau21601	Novel	hum
34	31	70.5	203	7	ADC46242	Adc46242	Human	neo
35	31	70.5	222	8	ADS44574	Ads44574	Bacterial	
36	31	70.5	240	6	ABP80613	Abp80613	N. gonorr	
37	31	70.5	343	4	ABB60195	Abb60195	Drosophil	
38	31	70.5	394	4	ABB70391	Abb70391	Drosophil	
39	31	70.5	460	5	ABP27238	Abp27238	Streptoco	
40	31	70.5	514	7	ADF03970	Adf03970	Bacterial	
41	31	70.5	534	8	ADJ76282	Adj76282	Marker	ge
42	31	70.5	752	8	ADS25075	Ads25075	Bacterial	
43	31	70.5	791	6	ABR53460	Abr53460	Protein	s
44	31	70.5	791	7	ADK64604	Adk64604	Disease	t
45	31	70.5	796	5	ABP73364	Abp73364	Candida	a

#### ALIGNMENTS

##### RESULT 1

ABB81970

ID ABB81970 standard; peptide; 9 AA.

XX

AC ABB81970;

XX

DT 25-NOV-2002 (first entry)

XX

DE 30 kDa ragweed pollen allergen tryptic peptide 3.

XX

KW Ragweed; pollen; allergen; Ambt 7; glycoprotein; antiallergic;  
KW immunotherapy; disulphide protein.

XX

OS Ambrosia elatior.

XX

FH Key Location/Qualifiers

FT Misc-difference 1  
 FT /label= Leu or Ile  
 XX  
 PN WO200263012-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 04-FEB-2002; 2002WO-US003346.  
 XX  
 PR 05-FEB-2001; 2001US-0266686P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Buchanan BB, Del Val G, Frick OL;  
 XX  
 DR WPI; 2002-657539/70.  
 XX  
 PT New ragweed pollen allergens, useful in allergy testing and immunotherapy  
 PT regimens, particularly for treating sensitivity to pollen or pollen  
 PT allergy (e.g. anaphylaxis, or symptoms of hives or asthma) in a mammal,  
 PT especially a human.  
 XX  
 PS Claim 1; Page 53; 70pp; English.  
 XX  
 CC The invention relates to an isolated pollen allergen purified from  
 CC ragweed pollen, substantially free of any other pollen proteins, or a  
 CC protein that is an antigenic fragment of a pollen allergen Ambt 7. The  
 CC allergen is characterized by the following physiochemical and biological  
 CC properties: (a) being contained in pollen extracts; (b) a glycoprotein;  
 CC (c) a sulphydryl group containing protein; (d) a molecular weight of  
 CC about 30 kDa as determined by SDS-polyacrylamide gel electrophoresis; and  
 CC (e) possessing allergen activity. The pollen allergen, or antigenic  
 CC protein fragment of the pollen allergen Ambt 7, or composition is useful  
 CC for treating sensitivity to pollen or pollen allergy in a mammal. This  
 CC allergy includes anaphylaxis or atopy, which includes the symptoms of hay  
 CC fever, asthma or hives. The allergen is also useful in allergy testing  
 CC and immunotherapy regimens. Sequences ABB81968-978 represent tryptic  
 CC peptide fragments of the 30 kDa ragweed complete pollen extract  
 CC disulphide protein allergen  
 XX  
 SQ Sequence 9 AA;

Query Match 97.7%; Score 43; DB 5; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
 |||||  
 Db 2 YGLVQFNR 9

RESULT 2  
 AAU56543  
 ID AAU56543 standard; protein; 99 AA.  
 XX  
 AC AAU56543;  
 XX

DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #17439.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US012865.  
 XX  
 PR 21-APR-2000; 2000US-0199047P.  
 PR 02-JUN-2000; 2000US-0208841P.  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59577.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 XX  
 PS Example 1; SEQ ID NO 17738; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 99 AA;

Query Match 75.0%; Score 33; DB 4; Length 99;  
Best Local Similarity 62.5%; Pred. No. 47;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
|||::| |  
Db 36 YGLIEFTR 43

RESULT 3  
ABM53062

ID ABM53062 standard; protein; 99 AA.  
XX  
AC ABM53062;  
XX  
DT 20-OCT-2003 (first entry)  
XX  
DE Propionibacterium acnes predicted ORF-encoded polypeptide #17738.  
XX  
KW Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO2003033515-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 11-OCT-2002; 2002WO-US032727.  
XX  
PR 15-OCT-2001; 2001US-00978825.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieve-Douglass J;  
XX  
DR WPI; 2003-381789/36.  
DR N-PSDB; ACF64506.  
XX  
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
PS Example 1; SEQ ID NO 17738; 1481pp; English.  
XX  
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared

CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 99 AA;

Query Match 75.0%; Score 33; DB 6; Length 99;  
 Best Local Similarity 62.5%; Pred. No. 47;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
 |||::| |  
 Db 36 YGLIEFTR 43

#### RESULT 4

ABU28434

ID ABU28434 standard; protein; 170 AA.

XX

AC ABU28434;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #13961.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Enterobacter cloacae.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA32304.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 56358; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 170 AA;

Query Match 75.0%; Score 33; DB 6; Length 170;  
 Best Local Similarity 85.7%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
 |||||  
 Db 106 YGLVMFN 112



ADM26790

ID ADM26790 standard; protein; 192 AA.

XX

AC ADM26790;

XX

DT 20-MAY-2004 (first entry)

XX

DE Hyperthermophile Methanopyrus kandleri protein #1396.

XX

KW hyperthermophile; protein stability enhancement;  
KW protein activity enhancement.

XX

OS Methanopyrus kandleri.

XX

PN WO2003076575-A2.

XX

PD 18-SEP-2003.

XX

PF 04-MAR-2003; 2003WO-US006664.

XX

PR 04-MAR-2002; 2002US-0361742P.

PR 14-MAY-2002; 2002US-0380423P.

PR 16-SEP-2002; 2002US-0410974P.

XX

PA (FIDE-) FIDELITY SYSTEMS INC.

PA (MALY/) MALYKH A.

XX

PI Slesarev AI, Pavlov A, Pavlova N, Kozyavkin S;

XX

DR WPI; 2003-748383/70.

DR N-PSDB; ADM27081.

XX

PT New isolated nucleic acids encoding any of about 1700 Methanopyrus  
PT kandleri proteins, and the encoded proteins, useful as a medicaments or  
PT as diagnostic agents.

XX

PS Claim 31; SEQ ID NO 1396; 1023pp; English.

XX

CC The invention comprises the amino acid sequence of proteins from the  
CC hyperthermophile Methanopyrus kandleri, the invention also comprises the  
CC complete genome from Methanopyrus kandleri. The Methanopyrus kandleri  
CC proteins of the invention are useful for enhancing the stability and/or  
CC activity of other proteins. The Methanopyrus kandleri genome is useful in  
CC a variety of diagnostic and analytical methods. The present amino acid  
CC sequence represents a Methanopyrus kandleri protein of the invention.

XX

SQ Sequence 192 AA;

Query Match 75.0%; Score 33; DB 7; Length 192;

Best Local Similarity 75.0%; Pred. No. 96;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9

||||:|

Db 134 YGLVKFER 141

RESULT 6

ABP25912

ID ABP25912 standard; protein; 253 AA.

XX

AC ABP25912;

XX

DT 02-JUL-2002 (first entry)

XX

DE Streptococcus polypeptide SEQ ID NO 1000.

XX

KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX

OS Streptococcus agalactiae.

XX

PN WO200234771-A2.

XX

PD 02-MAY-2002.

XX

PF 29-OCT-2001; 2001WO-GB004789.

XX

PR 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

PR 07-MAR-2001; 2001GB-00005640.

XX

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX

PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;

PI Tettelin H;

XX

DR WPI; 2002-352536/38.

DR N-PSDB; ABN66543.

XX

PT New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.

XX

PS Claim 1; Page 3254; 4525pp; English.

XX

CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins  
XX  
SQ Sequence 253 AA;

Query Match 75.0%; Score 33; DB 5; Length 253;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
||| :||  
Db 7 YGLVLYNR 14

RESULT 7

AAB51794

ID AAB51794 standard; protein; 329 AA.

XX

AC AAB51794;

XX

DT 16-FEB-2001 (first entry)

XX

DE Gene 21 human secreted protein homologous amino acid sequence #123.

XX

KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW nootropic; neuroprotective; antibacterial; virucide; fungicide;  
KW ophthalmological; vulnerary; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorders; cancer; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;  
KW Alzheimer's disease; ocular disorder; wound healing; skin aging.

XX

OS Homo sapiens.

XX

PN WO200061625-A1.

XX

PD 19-OCT-2000.

XX

PF 06-APR-2000; 2000WO-US008981.

XX

PR 09-APR-1999; 99US-0128701P.

PR 20-JAN-2000; 2000US-0177166P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX

PI Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI; 2000-619226/59.

XX

PT New nucleic acid molecules encoding 48 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives.

XX

PS Disclosure; Page 37-38; 500pp; English.

XX

CC Polynucleotide sequences AAC93422 - AAC93449 represent cDNA encoding  
CC human secreted proteins AAB51724 - AAB51777. Sequences AAB51778 -

CC AAB51825 represent alternative polypeptides encoded by the genes, and  
 CC amino acid sequences to which they are homologous. The genes and proteins  
 CC have activities dependent on the tissues and cells in which they are  
 CC expressed. Examples of their activities include immunosuppressive;  
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;  
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;  
 CC virucide; fungicide; ophthalmological; and vulnerary. The secreted  
 CC proteins, polynucleotides, antagonists and agonists may be useful in  
 CC treating, preventing and/or diagnosing diseases and disorders such as  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,  
 CC infections caused by bacteria, viruses and fungi and ocular disorders  
 CC e.g. corneal infection. The polypeptides can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,  
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional  
 CC components. Oligonucleotide AAC93413 - AAC93421 and peptide AAB51723 are  
 CC used in the isolation and characterisation of the proteins and  
 CC polynucleotides of the invention  
 XX  
 SQ Sequence 329 AA;

Query Match 75.0%; Score 33; DB 3; Length 329;  
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YGLVQFN 8  
 |||||:  
 Db 91 YGLVQFS 97

# RESULT 8

ADA33363

ID ADA33363 standard; protein; 819 AA.

XX

AC ADA33363;

XX

DT 20-NOV-2003 (first entry)

XX

DE Acinetobacter baumannii protein #524.

XX

KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;  
 KW plant biocontrol agent.

XX

OS Acinetobacter baumannii.

XX

PN US6562958-B1.

XX

PD 13-MAY-2003.

XX

PF 04-JUN-1999; 99US-00328352.

XX

PR 09-JUN-1998; 98US-0088701P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Breton G, Bush D;

XX

DR WPI; 2003-576092/54.

DR N-PSDB; ADA29237.

XX

PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
PT for diagnosing a bacterial disease, as components of antibacterial  
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
PT plants.

XX

PS Example; SEQ ID NO 4650; 328pp; English.

XX

CC The invention relates to isolated Acinetobacter baumannii nucleic acids.  
CC The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other Acinetobacter species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents the amino acid sequence of an A.  
CC baumannii protein.

XX

SQ Sequence 819 AA;

Query Match 75.0%; Score 33; DB 6; Length 819;

Best Local Similarity 75.0%; Pred. No. 4.7e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9

||| |||

Db 258 YGYVDFNR 265

#### RESULT 9

ABB16566

ID ABB16566 standard; protein; 112 AA.

XX

AC ABB16566;

XX

DT 23-JAN-2002 (first entry)

XX

DE Human nervous system related polypeptide SEQ ID NO 5223.

XX

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX

OS Homo sapiens.

XX

PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.

PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0242221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.

PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-541565/60.

DR N-PSDB; ABA12892.

XX

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.

XX

PS Claim 11; SEQ ID NO 5223; 1701pp + Sequence Listing; English.

XX

CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of



CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 112 AA;

Query Match 72.7%; Score 32; DB 4; Length 112;  
Best Local Similarity 75.0%; Pred. No. 87;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
| :|||  
Db 67 YQVVQFNR 74

RESULT 10

AAAY36180

ID AAY36180 standard; protein; 124 AA.

XX

AC AAY36180;

XX

DT 23-SEP-1999 (first entry)

XX

DE Human secreted protein #52.

XX

KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;  
KW diagnostic; gene therapy; chromosome mapping; secretion vector.

XX

OS Homo sapiens.

XX

PN WO9925825-A2.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-IB001862.

XX

PR 13-NOV-1997; 97US-0066677P.

PR 17-DEC-1997; 97US-0069957P.

PR 09-FEB-1998; 98US-0074121P.

PR 13-APR-1998; 98US-0081563P.

PR 10-AUG-1998; 98US-0096116P.

PR 04-SEP-1998; 98US-0099273P.

XX

PA (GEST ) GENSET.

XX

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX

DR WPI; 1999-347472/29.

DR N-PSDB; AAX97864.

XX

PT Extended cDNAs encoding secreted proteins.

XX

PS Claim 7; Page 288-289; 307pp; English.

XX

CC AAY36129-Y36222 represent novel human secreted proteins encoded by the  
CC extended cDNA sequences represented in AAX97813-X97906. The proteins of  
CC the invention have cytostatic, thrombotic and osteopathic activity. The  
CC extended cDNAs can be used to express secreted proteins or parts of them

CC or to obtain antibodies capable of binding to the secreted proteins. They  
CC may also be used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. Uses also include design of expression vectors and  
CC secretion vectors  
XX  
SQ Sequence 124 AA;

Query Match 72.7%; Score 32; DB 2; Length 124;  
Best Local Similarity 75.0%; Pred. No. 97;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
| :|||||  
Db 78 YQVVQFNR 85

RESULT 11

AAY36133

ID AAY36133 standard; protein; 124 AA.

XX

AC AAY36133;

XX

DT 23-SEP-1999 (first entry)

XX

DE Human secreted protein #5.

XX

KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;  
KW diagnostic; gene therapy; chromosome mapping; secretion vector.

XX

OS Homo sapiens.

XX

PN WO9925825-A2.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-IB001862.

XX

PR 13-NOV-1997; 97US-0066677P.

PR 17-DEC-1997; 97US-0069957P.

PR 09-FEB-1998; 98US-0074121P.

PR 13-APR-1998; 98US-0081563P.

PR 10-AUG-1998; 98US-0096116P.

PR 04-SEP-1998; 98US-0099273P.

XX

PA (GEST ) GENSET.

XX

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX

DR WPI; 1999-347472/29.

DR N-PSDB; AAX97817.

XX

PT Extended cDNAs encoding secreted proteins.

XX

PS Example 28; Page 234; 307pp; English.

XX

CC AAY36129-Y36222 represent novel human secreted proteins encoded by the  
CC extended cDNA sequences represented in AAX97813-X97906. The proteins of

CC the invention have cytostatic, thrombotic and osteopathic activity. The  
CC extended cDNAs can be used to express secreted proteins or parts of them  
CC or to obtain antibodies capable of binding to the secreted proteins. They  
CC may also be used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. Uses also include design of expression vectors and  
CC secretion vectors

XX

SQ Sequence 124 AA;

Query Match 72.7%; Score 32; DB 2; Length 124;  
Best Local Similarity 75.0%; Pred. No. 97;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :|||||  
Db 78 YQVVQFNR 85

# RESULT 12

AAY36210

ID AAY36210 standard; protein; 124 AA.

XX

AC AAY36210;

XX

DT 23-SEP-1999 (first entry)

XX

DE Human secreted protein #82..

XX

KW Secreted protein; human; cytostatic; thrombotic; osteopathic; forensic;  
KW diagnostic; gene therapy; chromosome mapping; secretion vector.

XX

OS Homo sapiens.

XX

PN WO9925825-A2.

XX

PD 27-MAY-1999.

XX

PF 13-NOV-1998; 98WO-IB001862.

XX

PR 13-NOV-1997; 97US-0066677P.

PR 17-DEC-1997; 97US-0069957P.

PR 09-FEB-1998; 98US-0074121P.

PR 13-APR-1998; 98US-0081563P.

PR 10-AUG-1998; 98US-0096116P.

PR 04-SEP-1998; 98US-0099273P.

XX

PA (GEST ) GENSET.

XX

PI Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX

DR WPI; 1999-347472/29.

DR N-PSDB; AAX97894.

XX

PT Extended cDNAs encoding secreted proteins.

XX

PS Claim 7; Page 302-303; 307pp; English.

XX

CC AAY36129-Y36222 represent novel human secreted proteins encoded by the  
CC extended cDNA sequences represented in AAX97813-X97906. The proteins of  
CC the invention have cytostatic, thrombotic and osteopathic activity. The  
CC extended cDNAs can be used to express secreted proteins or parts of them  
CC or to obtain antibodies capable of binding to the secreted proteins. They  
CC may also be used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. Uses also include design of expression vectors and  
CC secretion vectors

XX

SQ Sequence 124 AA;

Query Match 72.7%; Score 32; DB 2; Length 124;  
Best Local Similarity 75.0%; Pred. No. 97;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :|||||  
Db 78 YQVVQFNR 85

#### RESULT 13

AAM39501

ID AAM39501 standard; protein; 124 AA.

XX

AC AAM39501;

XX

DT 22-OCT-2001 (first entry)

XX

DE Human polypeptide SEQ ID NO 2646.

XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.

XX

OS Homo sapiens.

XX

PN WO200153312-A1.

XX

PD 26-JUL-2001.

XX

PF 26-DEC-2000; 2000WO-US034263.

XX

PR 23-DEC-1999; 99US-00471275.

PR 21-JAN-2000; 2000US-00488725.

PR 25-APR-2000; 2000US-00552317.

PR 20-JUN-2000; 2000US-00598042.

PR 19-JUL-2000; 2000US-00620312.

PR 03-AUG-2000; 2000US-00653450.

PR 14-SEP-2000; 2000US-00662191.

PR 19-OCT-2000; 2000US-00693036.

PR 29-NOV-2000; 2000US-00727344.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;

XX

DR WPI; 2001-442253/47.

DR N-PSDB; AAI58657.

XX

PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.

XX

PS Example 4; SEQ ID NO 2646; 10078pp; English.

XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification

XX

SQ Sequence 124 AA;

Query Match 72.7%; Score 32; DB 4; Length 124;  
Best Local Similarity 75.0%; Pred. No. 97;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :|||||  
Db 78 YQVVQFNR 85

#### RESULT 14

ABB89828

ID ABB89828 standard; protein; 124 AA.

XX

AC ABB89828;

XX

DT 24-MAY-2002 (first entry)

XX

DE Human polypeptide SEQ ID NO 2204.

XX

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.

XX

OS Homo sapiens.

XX

PN WO200190304-A2.  
 XX  
 PD 29-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001WO-US016450.  
 XX  
 PR 19-MAY-2000; 2000US-0205515P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-122018/16.  
 DR N-PSDB; ABL90237.  
 XX  
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders.  
 XX  
 PS Claim 11; SEQ ID NO 2204; 2081pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 124 AA;

Query Match 72.7%; Score 32; DB 5; Length 124;  
 Best Local Similarity 75.0%; Pred. No. 97;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 | :|||  
 Db 78 YQVVQFNR 85

# RESULT 15

AAM52201

ID AAM52201 standard; protein; 124 AA.

XX

AC AAM52201;

XX

DT 08-FEB-2002 (first entry)  
 XX  
 DE Human MP-1 SEQ ID NO 3.  
 XX  
 KW Human; mouse; rat; antisense gene therapy; MP-1; MAP kinase Partner 1;  
 KW antiinflammatory; cytostatic; antimicrobial; infection; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 120  
 FT /label= unknown  
 FT /note= "Encoded by GNT"  
 XX  
 PN US6306606-B1.  
 XX  
 PD 23-OCT-2001.  
 XX  
 PF 22-NOV-2000; 2000US-00721822.  
 XX  
 PR 22-NOV-2000; 2000US-00721822.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 PA (UYVI-) UNIV VIRGINIA.  
 XX  
 PI Weber MJ, Wyatt J, Cowsert LM;  
 XX  
 DR WPI; 2002-040199/05.  
 DR N-PSDB; ABA83444.  
 XX  
 PT New antisense oligonucleotides for modulating the expression of MP-1 (MAP  
 PT kinase partner 1), for preventing, delaying or treating infection,  
 PT inflammation or tumor formation, especially in humans.  
 XX  
 PS Example 13; Col 47-48; 47pp; English.  
 XX  
 CC The invention relates to an antisense compound (ABA83459-ABA83576) which  
 CC is up to 30 nucleobases in length and that inhibits the expression of MP-  
 CC 1 (MAP kinase Partner 1) in cells or tissues comprising contacting the  
 CC cells or tissues in vitro with the antisense compound so that expression  
 CC of MP-1 is inhibited. The antisense compounds have potential  
 CC antiinflammatory, cytostatic and antimicrobial activity. The antisense  
 CC compounds are useful for diagnostics, therapeutics, prophylaxis or as  
 CC research reagents or kits. The antisense oligonucleotides are useful in  
 CC gene therapy for treating an animal, particularly a human, suspected of  
 CC having or being prone to a disease or condition associated with the  
 CC expression of MP-1. In particular, the antisense oligonucleotides are  
 CC useful for preventing, delaying or treating infection, inflammation or  
 CC tumour formation. The present sequence is that of a human MP-1  
 XX  
 SQ Sequence 124 AA;

Query Match 72.7%; Score 32; DB 5; Length 124;  
 Best Local Similarity 75.0%; Pred. No. 97;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9

Db 78 YOVVOFN R 85

Search completed: February 10, 2005, 15:48:39  
Job time : 80.4648 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

```
Run on:      February 10, 2005, 15:38:08 ; Search time 20.1549 Seconds
              (without alignments)
              33.334 Million cell updates/sec
```

```
Title:      US-10-067-484-3
Perfect score: 44
Sequence:   1 XYGLVQFNR 9
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :      Issued_Patents_AA:*
1:  /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2:  /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3:  /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4:  /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5:  /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6:  /cgn2_6/ptodata/1/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	% Query Match	Length	DB	ID	Description
1	33	75.0	819	4	US-09-328-352-4650	Sequence 4650, App
2	32	72.7	124	4	US-09-663-600A-91	Sequence 91, Appl
3	32	72.7	124	4	US-09-663-600A-185	Sequence 185, App
4	32	72.7	124	4	US-09-663-600A-215	Sequence 215, App
5	32	72.7	124	4	US-09-621-976-14	Sequence 14, Appl



6	32	72.7	296	4	US-09-492-709A-337	Sequence 337, App
7	31	70.5	194	4	US-09-270-767-33892	Sequence 33892, A
8	31	70.5	194	4	US-09-270-767-49109	Sequence 49109, A
9	31	70.5	303	4	US-09-248-796A-14416	Sequence 14416, A
10	31	70.5	514	4	US-09-543-681A-4255	Sequence 4255, Ap
11	30	68.2	126	4	US-09-232-290-35	Sequence 35, Appl
12	30	68.2	134	4	US-09-732-210-1742	Sequence 1742, Ap
13	30	68.2	242	4	US-09-270-767-47078	Sequence 47078, A
14	30	68.2	265	4	US-09-540-236-3285	Sequence 3285, Ap
15	30	68.2	266	4	US-09-270-767-31861	Sequence 31861, A
16	30	68.2	356	4	US-09-107-532A-4245	Sequence 4245, Ap
17	30	68.2	362	4	US-09-134-000C-3578	Sequence 3578, Ap
18	30	68.2	366	4	US-09-328-352-7292	Sequence 7292, Ap
19	30	68.2	377	4	US-09-107-532A-4318	Sequence 4318, Ap
20	30	68.2	399	4	US-09-543-681A-6125	Sequence 6125, Ap
21	30	68.2	402	3	US-09-134-001C-4138	Sequence 4138, Ap
22	30	68.2	735	3	US-08-539-205A-2	Sequence 2, Appli
23	30	68.2	735	4	US-09-392-163A-2	Sequence 2, Appli
24	30	68.2	755	4	US-09-107-532A-3693	Sequence 3693, Ap
25	30	68.2	1380	4	US-09-949-016-11688	Sequence 11688, A
26	30	68.2	1874	4	US-09-602-787A-46	Sequence 46, Appl
27	30	68.2	2777	4	US-10-220-587-4	Sequence 4, Appli
28	30	68.2	2780	4	US-10-220-587-2	Sequence 2, Appli
29	29	65.9	151	4	US-09-270-767-60568	Sequence 60568, A
30	29	65.9	168	4	US-09-270-767-42676	Sequence 42676, A
31	29	65.9	168	4	US-09-270-767-57995	Sequence 57995, A
32	29	65.9	192	4	US-09-248-796A-20050	Sequence 20050, A
33	29	65.9	243	4	US-09-252-991A-29870	Sequence 29870, A
34	29	65.9	253	4	US-09-270-767-41859	Sequence 41859, A
35	29	65.9	290	3	US-09-002-298-9	Sequence 9, Appli
36	29	65.9	290	3	US-09-058-489-8	Sequence 8, Appli
37	29	65.9	290	4	US-09-481-277-9	Sequence 9, Appli
38	29	65.9	508	4	US-09-270-767-45071	Sequence 45071, A
39	29	65.9	521	4	US-09-270-767-43965	Sequence 43965, A
40	29	65.9	527	2	US-08-535-276-3	Sequence 3, Appli
41	29	65.9	527	3	US-09-335-234-3	Sequence 3, Appli
42	29	65.9	559	4	US-09-583-110-3735	Sequence 3735, Ap
43	29	65.9	567	4	US-09-107-433-4592	Sequence 4592, Ap
44	29	65.9	1901	4	US-09-738-946-12	Sequence 12, Appl
45	28.5	64.8	254	4	US-09-270-767-38691	Sequence 38691, A

#### ALIGNMENTS

##### RESULT 1

US-09-328-352-4650

; Sequence 4650, Application US/09328352

; Patent No. 6562958

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4650  
; LENGTH: 819  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4650

Query Match 75.0%; Score 33; DB 4; Length 819;  
Best Local Similarity 75.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
|||  
Db 258 YGYVDFNR 265

RESULT 2

US-09-663-600A-91  
; Sequence 91, Application US/09663600A  
; Patent No. 6573068  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS  
; FILE REFERENCE: 31.US3.CIP  
; CURRENT APPLICATION NUMBER: US/09/663,600A  
; CURRENT FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 09/191,997  
; PRIOR FILING DATE: 1998-11-13  
; PRIOR APPLICATION NUMBER: 60/066,677  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/069,957  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/074,121  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/081,563  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/096,116  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: 60/099,273  
; PRIOR FILING DATE: 1998-09-04  
; NUMBER OF SEQ ID NOS: 229  
; SOFTWARE: Patent.pm  
; SEQ ID NO 91  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -97...-1  
US-09-663-600A-91

Query Match 72.7%; Score 32; DB 4; Length 124;  
Best Local Similarity 75.0%; Pred. No. 35;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy            2 YGLVQFNR 9  
              | :||||  
Db            78 YQVVQFNR 85

RESULT 3

US-09-663-600A-185  
; Sequence 185, Application US/09663600A  
; Patent No. 6573068  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueleret, Lydie  
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS  
; FILE REFERENCE: 31.US3.CIP  
; CURRENT APPLICATION NUMBER: US/09/663,600A  
; CURRENT FILING DATE: 2000-09-15  
; PRIOR APPLICATION NUMBER: 09/191,997  
; PRIOR FILING DATE: 1998-11-13  
; PRIOR APPLICATION NUMBER: 60/066,677  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: 60/069,957  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 60/074,121  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: 60/081,563  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/096,116  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: 60/099,273  
; PRIOR FILING DATE: 1998-09-04  
; NUMBER OF SEQ ID NOS: 229  
; SOFTWARE: Patent.pm  
; SEQ ID NO 185  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -97..-1  
US-09-663-600A-185

Query Match            72.7%;   Score 32;   DB 4;   Length 124;  
Best Local Similarity   75.0%;   Pred. No. 35;  
Matches       6;   Conservative    1;   Mismatches    1;   Indels       0;   Gaps       0;

Qy            2 YGLVQFNR 9  
              | :||||  
Db            78 YQVVQFNR 85

RESULT 4

US-09-663-600A-215  
; Sequence 215, Application US/09663600A  
; Patent No. 6573068  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste

```

; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS
; FILE REFERENCE: 31.US3.CIP
; CURRENT APPLICATION NUMBER: US/09/663,600A
; CURRENT FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/191,997
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/066,677
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/069,957
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: 1998-09-04
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: Patent.pm
; SEQ ID NO 215
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -97...-1
US-09-663-600A-215

```

```

Query Match          72.7%; Score 32; DB 4; Length 124;
Best Local Similarity 75.0%; Pred. No. 35;
Matches      6; Conservative      1; Mismatches      1; Indels      0; Gaps      0;

```

```

Qy      2 YGLVQFNR 9
        | :|||||
Db      78 YQVVQFNR 85

```

# RESULT 5

```

US-09-621-976-14
; Sequence 14, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 124
; TYPE: PRT

```

; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -97...-1  
US-09-621-976-14

Query Match 72.7%; Score 32; DB 4; Length 124;  
Best Local Similarity 75.0%; Pred. No. 35;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| : ||||  
Db 78 YQVVQFNR 85

RESULT 6

US-09-492-709A-337  
; Sequence 337, Application US/09492709A  
; Patent No. 6720139  
; GENERAL INFORMATION:  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Trawick, John  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Froelich, Jamie M.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; TITLE OF INVENTION: ESCHERICHIA COLI  
; FILE REFERENCE: ELITRA.001A  
; CURRENT APPLICATION NUMBER: US/09/492,709A  
; CURRENT FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 485  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 337  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: E. Coli  
US-09-492-709A-337

Query Match 72.7%; Score 32; DB 4; Length 296;  
Best Local Similarity 75.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
||| |||  
Db 49 YGLCQFGR 56

RESULT 7

US-09-270-767-33892  
; Sequence 33892, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 33892  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-33892

Query Match 70.5%; Score 31; DB 4; Length 194;  
Best Local Similarity 71.4%; Pred. No. 90;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
||:| ||  
Db 47 YGIVSFN 53

RESULT 8

US-09-270-767-49109  
; Sequence 49109, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49109  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-49109

Query Match 70.5%; Score 31; DB 4; Length 194;  
Best Local Similarity 71.4%; Pred. No. 90;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
||:| ||  
Db 47 YGIVSFN 53

RESULT 9

US-09-248-796A-14416  
; Sequence 14416, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132

```
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14416
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (14)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations
are unknown.
US-09-248-796A-14416
```

```
Query Match          70.5%; Score 31; DB 4; Length 303;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches      4; Conservative      4; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy      2 YGLVQFNR 9
        :|::|||:
Db      71 HGIIQFNQ 78
```

# RESULT 10

```
US-09-543-681A-4255
; Sequence 4255, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS
MIRABILIS FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4255
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4255
```

```
Query Match          70.5%; Score 31; DB 4; Length 514;
Best Local Similarity 71.4%; Pred. No. 2.5e+02;
Matches      5; Conservative      2; Mismatches      0; Indels      0; Gaps      0;
```

```
Qy      2 YGLVQFN 8
        |||::||
Db      276 YGLLRFN 282
```

RESULT 11

US-09-232-290-35

; Sequence 35, Application US/09232290A  
 ; Patent No. 6815540  
 ; GENERAL INFORMATION:  
 ; APPLICANT: PLUCKTHUN, ANDREAS  
 ; APPLICANT: NIEBA, LARS  
 ; APPLICANT: HONEGGER, ANNEMARIE  
 ; TITLE OF INVENTION: IMMUNOGLOBULIN SUPER FAMILY DOMAINS AND FRAGMENTS WITH  
 ; TITLE OF INVENTION: INCREASED SOLUBILITY  
 ; FILE REFERENCE: MORPHO/7  
 ; CURRENT APPLICATION NUMBER: US/09/232,290A  
 ; CURRENT FILING DATE: 1999-01-15  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/02230  
 ; EARLIER FILING DATE: 1996-05-23  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 35  
 ; LENGTH: 126  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

US-09-232-290-35

Query Match 68.2%; Score 30; DB 4; Length 126;  
 Best Local Similarity 85.7%; Pred. No. 91;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GLVQFNR 9  
 |||||  
 Db 10 GLVQFGR 16

RESULT 12

US-09-732-210-1742

; Sequence 1742, Application US/09732210  
 ; Patent No. 6573361  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bunkers, Greg J.  
 ; APPLICANT: Liang, Jihong  
 ; APPLICANT: Mittanck, Cindy A.  
 ; APPLICANT: Seale, Jeffrey W.  
 ; APPLICANT: Wu, Yonnie S.  
 ; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
 ; FILE REFERENCE: 38-21(15036)B  
 ; CURRENT APPLICATION NUMBER: US/09/732,210  
 ; CURRENT FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,513  
 ; PRIOR FILING DATE: 1999-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,340  
 ; PRIOR FILING DATE: 1999-12-07  
 ; NUMBER OF SEQ ID NOS: 1753  
 ; SEQ ID NO 1742  
 ; LENGTH: 134  
 ; TYPE: PRT  
 ; ORGANISM: Schizosaccharomyces pombe

US-09-732-210-1742



Query Match 68.2%; Score 30; DB 4; Length 134;  
Best Local Similarity 62.5%; Pred. No. 97;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
:|:||||  
Db 25 FGGIQFNR 32

RESULT 13

US-09-270-767-47078

; Sequence 47078, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47078  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: *Drosophila melanogaster*  
US-09-270-767-47078

Query Match 68.2%; Score 30; DB 4; Length 242;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GLVQFNR 9  
|||||  
Db 171 GLVQFRR 177

RESULT 14

US-09-540-236-3285

; Sequence 3285, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 3285  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: *M. catarrhalis*  
US-09-540-236-3285

Query Match 68.2%; Score 30; DB 4; Length 265;  
Best Local Similarity 71.4%; Pred. No. 2e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
||| :||  
Db 25 YGLAKFN 31

RESULT 15

US-09-270-767-31861  
; Sequence 31861, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 31861  
; LENGTH: 266  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-31861

Query Match 68.2%; Score 30; DB 4; Length 266;  
Best Local Similarity 85.7%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GLVQFNR 9  
|||||  
Db 195 GLVQFRR 201

Search completed: February 10, 2005, 16:02:07  
Job time : 21.1549 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:49:10 ; Search time 53.8732 Seconds  
(without alignments)  
54.586 Million cell updates/sec

Title: US-10-067-484-3  
Perfect score: 44  
Sequence: 1 XYGLVQFNR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Match	Length	DB	ID	Description	
1	43	97.7	9	14	US-10-067-484-3	Sequence 3, Appli	
2	43	97.7	9	14	US-10-067-620-3	Sequence 3, Appli	
3	33	75.0	47	16	US-10-437-963-121093	Sequence 121093,	
4	33	75.0	170	15	US-10-282-122A-56358	Sequence 56358, A	
5	33	75.0	189	15	US-10-424-599-144524	Sequence 144524,	
6	33	75.0	656	16	US-10-767-701-46396	Sequence 46396, A	
7	32	72.7	82	16	US-10-437-963-158329	Sequence 158329,	
8	32	72.7	97	16	US-10-437-963-200385	Sequence 200385,	
9	32	72.7	124	11	US-09-978-360A-697	Sequence 697, App	
10	32	72.7	124	11	US-09-978-360A-727	Sequence 727, App	
11	32	72.7	124	14	US-10-319-763-91	Sequence 91, Appl	
12	32	72.7	124	14	US-10-319-763-185	Sequence 185, App	
13	32	72.7	124	14	US-10-319-763-215	Sequence 215, App	
14	32	72.7	124	15	US-10-264-237-2204	Sequence 2204, Ap	
15	32	72.7	294	15	US-10-369-493-19705	Sequence 19705, A	
16	32	72.7	296	9	US-09-912-020-337	Sequence 337, App	

17	32	72.7	296	15	US-10-282-122A-42616	Sequence 42616, A
18	32	72.7	323	15	US-10-335-977-5007	Sequence 5007, Ap
19	32	72.7	425	16	US-10-437-963-158681	Sequence 158681,
20	32	72.7	444	14	US-10-156-761-13243	Sequence 13243, A
21	32	72.7	450	16	US-10-437-963-136461	Sequence 136461,
22	31	70.5	74	15	US-10-424-599-203155	Sequence 203155,
23	31	70.5	113	15	US-10-424-599-240610	Sequence 240610,
24	31	70.5	203	14	US-10-103-313-328	Sequence 328, App
25	31	70.5	222	15	US-10-369-493-23004	Sequence 23004, A
26	31	70.5	374	15	US-10-424-599-240614	Sequence 240614,
27	31	70.5	380	16	US-10-437-963-167122	Sequence 167122,
28	31	70.5	428	13	US-10-062-254-372	Sequence 372, App
29	31	70.5	438	15	US-10-424-599-265659	Sequence 265659,
30	31	70.5	752	15	US-10-369-493-14108	Sequence 14108, A
31	31	70.5	796	14	US-10-032-585-7201	Sequence 7201, Ap
32	30	68.2	60	16	US-10-437-963-143821	Sequence 143821,
33	30	68.2	109	16	US-10-437-963-176989	Sequence 176989,
34	30	68.2	110	15	US-10-424-599-206728	Sequence 206728,
35	30	68.2	119	16	US-10-437-963-177294	Sequence 177294,
36	30	68.2	205	15	US-10-424-599-168733	Sequence 168733,
37	30	68.2	274	16	US-10-437-963-203079	Sequence 203079,
38	30	68.2	297	15	US-10-425-114-46208	Sequence 46208, A
39	30	68.2	304	15	US-10-399-240-4	Sequence 4, Appli
40	30	68.2	313	16	US-10-767-701-41163	Sequence 41163, A
41	30	68.2	373	15	US-10-282-122A-47890	Sequence 47890, A
42	30	68.2	404	16	US-10-437-963-178788	Sequence 178788,
43	30	68.2	409	13	US-10-062-254-276	Sequence 276, App
44	30	68.2	409	15	US-10-424-599-251449	Sequence 251449,
45	30	68.2	410	15	US-10-424-599-251447	Sequence 251447,

#### ALIGNMENTS

##### RESULT 1

US-10-067-484-3

```
; Sequence 3, Application US/10067484
; Publication No. US20030170763A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; TITLE OF INVENTION: RAGWEED ALLERGENS
; FILE REFERENCE: 416272000200
; CURRENT APPLICATION NUMBER: US/10/067,484
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Ragweed
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 1
```

; OTHER INFORMATION: Xaa= Leucine or Isoleucine  
US-10-067-484-3

Query Match 97.7%; Score 43; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
|||||||  
Db 2 YGLVQFNR 9

RESULT 2

US-10-067-620-3

; Sequence 3, Application US/10067620  
; Publication No. US20030180225A1  
; GENERAL INFORMATION:  
; APPLICANT: Buchanan, Bob B.  
; APPLICANT: del Val, Gregorio  
; APPLICANT: Frick, Oscar L.  
; APPLICANT: Teuber, Suzanne S.  
; TITLE OF INVENTION: WALNUT AND RYEGRASS ALLERGENS  
; FILE REFERENCE: 416272003400  
; CURRENT APPLICATION NUMBER: US/10/067,620  
; CURRENT FILING DATE: 2002-02-04  
; PRIOR APPLICATION NUMBER: US 60/266,686  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 9  
; TYPE: PRT  
; ORGANISM: Ragweed  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 1  
; OTHER INFORMATION: Xaa= Leucine or Isoleucine  
US-10-067-620-3

Query Match 97.7%; Score 43; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
|||||||  
Db 2 YGLVQFNR 9

RESULT 3

US-10-437-963-121093

; Sequence 121093, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 121093  
; LENGTH: 47  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(47)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_24151C.1.pep  
US-10-437-963-121093

Query Match 75.0%; Score 33; DB 16; Length 47;  
Best Local Similarity 75.0%; Pred. No. 18;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
||| |:  
Db 27 YGLVSFHR 34

#### RESULT 4

US-10-282-122A-56358  
; Sequence 56358, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56358
;   LENGTH: 170
;   TYPE: PRT
;   ORGANISM: Enterobacter cloacae
US-10-232-122A-56358

```

```

Query Match          75.0%; Score 33; DB 15; Length 170;
Best Local Similarity 85.7%; Pred. No. 69;
Matches      6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2 YGLVQFN 8
        |||| ||
Db     106 YGLVMFN 112

```

# RESULT 5

```

US-10-424-599-144524
; Sequence 144524, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 144524
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_101518C.1.pep
US-10-424-599-144524

```

Query Match 75.0%; Score 33; DB 15; Length 189;  
Best Local Similarity 85.7%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
||| ||  
Db 103 YGLVNFN 109

RESULT 6

US-10-767-701-46396

; Sequence 46396, Application US/10767701

; Publication No. US20040172684A1

; GENERAL INFORMATION:

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement

; FILE REFERENCE: 38-21(53535)B

; CURRENT APPLICATION NUMBER: US/10/767,701

; CURRENT FILING DATE: 2004-01-29

; NUMBER OF SEQ ID NOS: 63128

; SEQ ID NO 46396

; LENGTH: 656

; TYPE: PRT

; ORGANISM: Sorghum bicolor

; FEATURE:

; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C3580\_1.pep

US-10-767-701-46396

Query Match 75.0%; Score 33; DB 16; Length 656;  
Best Local Similarity 75.0%; Pred. No. 2.8e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
|| ||| |  
Db 162 YGFVQFER 169

RESULT 7

US-10-437-963-158329

; Sequence 158329, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With



; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 158329  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_57814C.1.pep  
US-10-437-963-158329

Query Match 72.7%; Score 32; DB 16; Length 82;  
Best Local Similarity 62.5%; Pred. No. 53;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
|| :||:|  
Db 68 YGAIQFSR 75

RESULT 8

US-10-437-963-200385  
; Sequence 200385, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 200385  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_9585C.1.pep  
US-10-437-963-200385

Query Match 72.7%; Score 32; DB 16; Length 97;  
Best Local Similarity 85.7%; Pred. No. 63;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
||| |||  
Db 73 YGLQQFN 79

RESULT 9

US-09-978-360A-697

; Sequence 697, Application US/09978360A  
 ; Publication No. US20040110939A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edwards, Jean-Baptiste Dumas Milne  
 ; APPLICANT: Duclert, Aymeric  
 ; APPLICANT: Bougueleret, Lydie  
 ; APPLICANT: Jobert, Severin  
 ; APPLICANT: Clusel, Catherine  
 ; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides

; FILE REFERENCE: 56.US4.CIP  
 ; CURRENT APPLICATION NUMBER: US/09/978,360A  
 ; CURRENT FILING DATE: 2001-10-15  
 ; PRIOR APPLICATION NUMBER: US 60/066,677  
 ; PRIOR FILING DATE: 1997-11-13  
 ; PRIOR APPLICATION NUMBER: US 60/069,957  
 ; PRIOR FILING DATE: 1997-12-17  
 ; PRIOR APPLICATION NUMBER: US 60/074,121  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: US 60/081,563  
 ; PRIOR FILING DATE: 1998-04-13  
 ; PRIOR APPLICATION NUMBER: US 60/096,116  
 ; PRIOR FILING DATE: 1998-08-10  
 ; PRIOR APPLICATION NUMBER: US 60/099,273  
 ; PRIOR FILING DATE: -09-04  
 ; PRIOR APPLICATION NUMBER: US 09/191,997  
 ; PRIOR FILING DATE: 1998-11-13  
 ; PRIOR APPLICATION NUMBER: US 09/215,435  
 ; PRIOR FILING DATE: 1998-12-17  
 ; PRIOR APPLICATION NUMBER: PCT/IB98/02122  
 ; PRIOR FILING DATE: 1998-12-17  
 ; PRIOR APPLICATION NUMBER: US 09/247,155  
 ; PRIOR FILING DATE: 1999-02-09  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 810  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 697  
 ; LENGTH: 124  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SIGNAL  
 ; LOCATION: -97...-1

US-09-978-360A-697

Query Match 72.7%; Score 32; DB 11; Length 124;  
 Best Local Similarity 75.0%; Pred. No. 81;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 | :||||  
 Db 78 YQVVQFNR 85

RESULT 10

US-09-978-360A-727

; Sequence 727, Application US/09978360A  
; Publication No. US20040110939A1  
; GENERAL INFORMATION:  
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne  
; APPLICANT: Duclert, Aymeric  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Jobert, Severin  
; APPLICANT: Clusel, Catherine  
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal  
Peptides  
; FILE REFERENCE: 56.US4.CIP  
; CURRENT APPLICATION NUMBER: US/09/978,360A  
; CURRENT FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: US 60/066,677  
; PRIOR FILING DATE: 1997-11-13  
; PRIOR APPLICATION NUMBER: US 60/069,957  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: US 60/074,121  
; PRIOR FILING DATE: 1998-02-09  
; PRIOR APPLICATION NUMBER: US 60/081,563  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: US 60/096,116  
; PRIOR FILING DATE: 1998-08-10  
; PRIOR APPLICATION NUMBER: US 60/099,273  
; PRIOR FILING DATE: -09-04  
; PRIOR APPLICATION NUMBER: US 09/191,997  
; PRIOR FILING DATE: 1998-11-13  
; PRIOR APPLICATION NUMBER: US 09/215,435  
; PRIOR FILING DATE: 1998-12-17  
; PRIOR APPLICATION NUMBER: PCT/IB98/02122  
; PRIOR FILING DATE: 1998-12-17  
; PRIOR APPLICATION NUMBER: US 09/247,155  
; PRIOR FILING DATE: 1999-02-09  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 810  
; SOFTWARE: Patent.pm  
; SEQ ID NO 727  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -97...-1  
US-09-978-360A-727

Query Match 72.7%; Score 32; DB 11; Length 124;  
Best Local Similarity 75.0%; Pred. No. 81;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :|||||  
Db 78 YQVVQFNR 85

RESULT 11

US-10-319-763-91

; Sequence 91, Application US/10319763  
 ; Publication No. US20030144490A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
 ; APPLICANT: Duclert, Aymeric  
 ; APPLICANT: Bougueleret, Lydie  
 ; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS  
 ; FILE REFERENCE: G-031.US04.DIV  
 ; CURRENT APPLICATION NUMBER: US/10/319,763  
 ; CURRENT FILING DATE: 2002-12-10  
 ; PRIOR APPLICATION NUMBER: 60/066,677  
 ; PRIOR FILING DATE: 1997-11-13  
 ; PRIOR APPLICATION NUMBER: 60/069,957  
 ; PRIOR FILING DATE: 1997-12-17  
 ; PRIOR APPLICATION NUMBER: 60/074,121  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: 60/081,563  
 ; PRIOR FILING DATE: 1998-04-13  
 ; PRIOR APPLICATION NUMBER: 60/096,116  
 ; PRIOR FILING DATE: 1998-08-10  
 ; PRIOR APPLICATION NUMBER: 60/099,273  
 ; PRIOR FILING DATE: 1998-09-04  
 ; NUMBER OF SEQ ID NOS: 229  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 91  
 ; LENGTH: 124  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SIGNAL  
 ; LOCATION: -97...-1

US-10-319-763-91

Query Match 72.7%; Score 32; DB 14; Length 124;  
 Best Local Similarity 75.0%; Pred. No. 81;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
 | :|||||  
 Db 78 YQVVQFNR 85

RESULT 12

US-10-319-763-185

; Sequence 185, Application US/10319763  
 ; Publication No. US20030144490A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
 ; APPLICANT: Duclert, Aymeric  
 ; APPLICANT: Bougueleret, Lydie  
 ; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS  
 ; FILE REFERENCE: G-031.US04.DIV  
 ; CURRENT APPLICATION NUMBER: US/10/319,763  
 ; CURRENT FILING DATE: 2002-12-10  
 ; PRIOR APPLICATION NUMBER: 60/066,677

; PRIOR FILING DATE: 1997-11-13  
 ; PRIOR APPLICATION NUMBER: 60/069,957  
 ; PRIOR FILING DATE: 1997-12-17  
 ; PRIOR APPLICATION NUMBER: 60/074,121  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: 60/081,563  
 ; PRIOR FILING DATE: 1998-04-13  
 ; PRIOR APPLICATION NUMBER: 60/096,116  
 ; PRIOR FILING DATE: 1998-08-10  
 ; PRIOR APPLICATION NUMBER: 60/099,273  
 ; PRIOR FILING DATE: 1998-09-04  
 ; NUMBER OF SEQ ID NOS: 229  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 185  
 ; LENGTH: 124  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SIGNAL  
 ; LOCATION: -97...-1  
 US-10-319-763-185

Query Match 72.7%; Score 32; DB 14; Length 124;  
 Best Local Similarity 75.0%; Pred. No. 81;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
 | :|||||  
 Db 78 YQVVQFNR 85

# RESULT 13

US-10-319-763-215

; Sequence 215, Application US/10319763  
 ; Publication No. US20030144490A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, Jean-Baptiste  
 ; APPLICANT: Duclert, Aymeric  
 ; APPLICANT: Bougueleret, Lydie  
 ; TITLE OF INVENTION: EXTENDED CDNAS FOR SECRETED PROTEINS  
 ; FILE REFERENCE: G-031.US04.DIV  
 ; CURRENT APPLICATION NUMBER: US/10/319,763  
 ; CURRENT FILING DATE: 2002-12-10  
 ; PRIOR APPLICATION NUMBER: 60/066,677  
 ; PRIOR FILING DATE: 1997-11-13  
 ; PRIOR APPLICATION NUMBER: 60/069,957  
 ; PRIOR FILING DATE: 1997-12-17  
 ; PRIOR APPLICATION NUMBER: 60/074,121  
 ; PRIOR FILING DATE: 1998-02-09  
 ; PRIOR APPLICATION NUMBER: 60/081,563  
 ; PRIOR FILING DATE: 1998-04-13  
 ; PRIOR APPLICATION NUMBER: 60/096,116  
 ; PRIOR FILING DATE: 1998-08-10  
 ; PRIOR APPLICATION NUMBER: 60/099,273  
 ; PRIOR FILING DATE: 1998-09-04  
 ; NUMBER OF SEQ ID NOS: 229  
 ; SOFTWARE: Patent.pm

; SEQ ID NO 215  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: -97...-1

US-10-319-763-215

Query Match 72.7%; Score 32; DB 14; Length 124;  
Best Local Similarity 75.0%; Pred. No. 81;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :||||  
Db 78 YQVVQFNR 85

RESULT 14

US-10-264-237-2204

; Sequence 2204, Application US/10264237  
; Publication No. US20040009491A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA131P1  
; CURRENT APPLICATION NUMBER: US/10/264,237  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/16450  
; PRIOR FILING DATE: 2001-05-18  
; PRIOR APPLICATION NUMBER: US 60/205,515  
; PRIOR FILING DATE: 2000-05-19  
; NUMBER OF SEQ ID NOS: 2876  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 2204  
; LENGTH: 124  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-264-237-2204

Query Match 72.7%; Score 32; DB 15; Length 124;  
Best Local Similarity 75.0%; Pred. No. 81;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :||||  
Db 78 YQVVQFNR 85

RESULT 15

US-10-369-493-19705

; Sequence 19705, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
 OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 19705  
 ; LENGTH: 294  
 ; TYPE: PRT  
 ; ORGANISM: Nitrosomonas europaea  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(294)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 US-10-369-493-19705

Query Match 72.7%; Score 32; DB 15; Length 294;  
 Best Local Similarity 71.4%; Pred. No. 2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
 ||:||  
 Db 145 YGVVEFN 151

Search completed: February 10, 2005, 16:41:30  
 Job time : 53.8732 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 13.9437 Seconds  
 (without alignments)  
 62.104 Million cell updates/sec

Title: US-10-067-484-3  
 Perfect score: 44  
 Sequence: 1 XYGLVQFNR 9

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length					
1	33	75.0	63	2	D69115			hypothetical prote	
2	33	75.0	650	2	T32897			hypothetical prote	
3	33	75.0	662	2	T47649			ABC transporter-li	
4	32	72.7	296	2	H90896			hypothetical prote	
5	32	72.7	296	2	G85720			hypothetical prote	
6	32	72.7	296	2	B64908			ydeH protein - Esc	
7	32	72.7	417	2	C69795			glutamyl-tRNA (Gln)	
8	32	72.7	476	2	T51583			glutamyl-tRNA (Gln)	
9	32	72.7	476	2	T44293			glutamyl-tRNA (Gln)	
10	32	72.7	751	2	D71860			probable outer mem	
11	31	70.5	152	2	E64842			probable monooxyge	
12	31	70.5	152	2	E90785			probable 4-hydroxy	
13	31	70.5	152	2	C85645			probable 4-hydroxy	
14	31	70.5	345	2	S72490			N-acetyl-gamma-glu	
15	31	70.5	364	2	A70474			conserved hypothet	
16	31	70.5	428	2	T03934			DNA binding protei	
17	31	70.5	482	2	T01932			RNA binding protei	
18	31	70.5	570	2	C86440			PPR-repeat protein	
19	31	70.5	791	2	S61698			hypothetical prote	
20	31	70.5	794	1	S37606			SEC18 protein - ye	
21	31	70.5	3655	2	T38084			TRAP-like protein	
22	30	68.2	47	2	F81235			hypothetical prote	
23	30	68.2	81	2	T06548			RNA-directed DNA p	
24	30	68.2	88	2	B47760			retrovirus-related	
25	30	68.2	123	2	T43380			ribosomal protein	
26	30	68.2	133	2	G91092			hypothetical prote	
27	30	68.2	134	2	T37749			60s ribosomal prot	
28	30	68.2	136	2	T19479			hypothetical prote	
29	30	68.2	167	2	C85938			hypothetical prote	
30	30	68.2	207	2	T03331			gene e36 protein -	
31	30	68.2	223	2	H87335			conserved hypothet	
32	30	68.2	229	2	A81169			hypothetical prote	
33	30	68.2	231	2	G81936			hypothetical prote	
34	30	68.2	292	2	A97512			glucose 1-dehydrog	
35	30	68.2	296	2	D75597			glucose-1-phosphat	
36	30	68.2	297	2	AC1706			hypothetical prote	
37	30	68.2	297	2	AD1335			hypothetical prote	
38	30	68.2	304	2	S24263			seed storage prote	



39	30	68.2	357	2	G86906	hypothetical prote
40	30	68.2	400	2	G83790	aminotransferase B
41	30	68.2	404	2	F71718	alanine racemase (
42	30	68.2	408	2	C70379	hypothetical prote
43	30	68.2	449	2	T48511	aspartate transami
44	30	68.2	468	2	T33516	hypothetical prote
45	30	68.2	474	2	AB2161	hypothetical prote

# ALIGNMENTS

## RESULT 1

D69115

hypothetical protein MTH1857 - Methanobacterium thermoautotrophicum (strain Delta H)

C;Species: Methanobacterium thermoautotrophicum

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C;Accession: D69115

R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.; Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.; Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; Caruso, A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.; McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.

J. Bacteriol. 179, 7135-7155, 1997

A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional analysis and comparative genomics.

A;Reference number: A69000; MUID:98037514; PMID:9371463

A;Accession: D69115

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-63 <MTH>

A;Cross-references: UNIPROT:O27885; GB:AE000938; GB:AE000666; NID:g2622986;

PIDN:AAB86323.1; PID:g2622993

A;Experimental source: strain Delta H

C;Genetics:

A;Gene: MTH1857

Query Match 75.0%; Score 33; DB 2; Length 63;  
 Best Local Similarity 75.0%; Pred. No. 3.8;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 |||| :||  
 Db 27 YGLVNWNR 34

## RESULT 2

T32897

hypothetical protein C42C1.10 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T32897

R;Murray, J.; Rohlfing, T.; Davidson, S.

submitted to the EMBL Data Library, January 1998

A;Description: The sequence of *C. elegans* cosmid C42C1.  
A;Reference number: Z21243  
A;Accession: T32897  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-650 <MUR>  
A;Cross-references: EMBL:AF043695; PIDN:AAB97551.1; GSPDB:GN00019; CESP:C42C1.10  
A;Experimental source: strain Bristol N2; clone C42C1  
C;Genetics:  
A;Gene: CESP:C42C1.10  
A;Map position: 1  
A;Introns: 59/3; 136/3; 228/3; 293/1; 359/1; 371/3; 597/1

Query Match 75.0%; Score 33; DB 2; Length 650;  
Best Local Similarity 85.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
|||||:  
Db 91 YGLVQFS 97

#### RESULT 3

T47649

ABC transporter-like protein - *Arabidopsis thaliana*

N;Alternate names: protein T15C9.100

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004

C;Accession: T47649

R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24470

A;Accession: T47649

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-662 <MEW>

A;Cross-references: UNIPROT:Q9M2V6; EMBL:AL132970

A;Experimental source: cultivar Columbia; BAC clone T15C9

C;Genetics:

A;Map position: 3

A;Note: T15C9.100

C;Superfamily: *Arabidopsis thaliana* probable ATP-binding cassette protein  
F12L6.1; ATP-binding cassette homology

Query Match 75.0%; Score 33; DB 2; Length 662;  
Best Local Similarity 85.7%; Pred. No. 44;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GLVQFNR 9  
|||:|||  
Db 300 GLVEFNR 306

#### RESULT 4

H90896

hypothetical protein ECs2144 [imported] - *Escherichia coli* (strain O157:H7,  
substrain RIMD 0509952)

C;Species: Escherichia coli  
 C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
 C;Accession: H90896  
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.  
 A;Reference number: A99629; MUID:21156231; PMID:11258796  
 A;Accession: H90896  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-296 <HAY>  
 A;Cross-references: UNIPROT:Q8XB26; GB:BA000007; PIDN:BA35567.1; PID:g13361610; GSPDB:GN00154  
 A;Experimental source: strain O157:H7, substrain RIMD 0509952  
 C;Genetics:  
 A;Gene: ECs2144

Query Match 72.7%; Score 32; DB 2; Length 296;  
 Best Local Similarity 75.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 ||| |||  
 Db 49 YGLCQFGR 56

# RESULT 5

G85720

hypothetical protein ydeH [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C;Species: Escherichia coli  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C;Accession: G85720  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamosis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: G85720  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-296 <STO>  
 A;Cross-references: UNIPROT:Q8XB26; GB:AE005174; NID:g12515121; PIDN:AAG56227.1; GSPDB:GN00145; UWGP:Z2163  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: ydeH

Query Match 72.7%; Score 32; DB 2; Length 296;  
 Best Local Similarity 75.0%; Pred. No. 32;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
||| |||  
Db 49 YGLCQFGR 56

RESULT 6

B64908

ydeH protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004

C;Accession: B64908

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: B64908

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-296 <BLAT>

A;Cross-references: UNIPROT:P31129; GB:AE000251; GB:U00096; NID:g1787814;

PIDN:AAC74608.1; PID:g1787816; UWGP:b1535

A;Experimental source: strain K-12, substrain MG1655

C;Genetics:

A;Gene: ydeH

Query Match 72.7%; Score 32; DB 2; Length 296;  
Best Local Similarity 75.0%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
||| |||  
Db 49 YGLCQFGR 56

RESULT 7

C69795

glutamyl-tRNA(Gln) amidotransferase (EC 2.6.-.-) chain B [validated] - Bacillus subtilis

C;Species: Bacillus subtilis

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 18-Aug-2000

C;Accession: C69795

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.;

Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: C69795

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-417 <KUN>

A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12489.1; PID:g2632983

A;Experimental source: strain 168

C;Genetics:

A;Gene: gatB; yerN

A;Note: transcription unit gatCAB

C;Complex: heterotrimer; consists of chain A (PIR:B69795), chain B (PIR:C69795), and chain C (PIR:A69795) [validated, MUID:98004482]

C;Function:

A;Description: (EC 2.6.-.-); glutamyl-tRNA(Gln) amidotransferase [validated, MUID:98004482]; converts misacylated Glu-tRNA(Gln) to correctly charged Gln-tRNA(Gln) by transamidation

A;Pathway: Gln-tRNA(Gln) biosynthesis

A;Note: tRNA-dependent amidation of mischarged Glu-tRNA(Gln) is the only pathway for the synthesis of Gln-tRNA(Gln) in *Bacillus subtilis* and several other species

C;Superfamily: PET112 protein

C;Keywords: aminotransferase; ATP

Query Match 72.7%; Score 32; DB 2; Length 417;  
Best Local Similarity 75.0%; Pred. No. 45;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| | | |  
Db 78 YSLVDFNR 85

RESULT 8

T51583

glutamyl-tRNA (Gln) amidotransferase (EC 2.6.-.-) chain B [validated] - *Bacillus subtilis*

C;Species: *Bacillus subtilis*  
 C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 02-Sep-2000  
 C;Accession: T51583  
 R;Curnow, A.W.; Hong, K.W.; Yuan, R.; Kim, S.I.; Martins, O.; Winkler, W.;  
 Henkin, T.M.; Soll, D.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 11819-11826, 1997  
 A;Title: Glu-tRNAGln amidotransferase: A novel heterotrimeric enzyme required  
 for correct decoding of glutamine codons during translation.  
 A;Reference number: Z25395; MUID:98004482; PMID:9342321  
 A;Accession: T51583  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-476 <CUR>  
 A;Cross-references: EMBL:AF008553; PIDN:AAB83965.1  
 A;Experimental source: strain 168  
 C;Genetics:  
 A;Gene: gatB  
 C;Function:  
 A;Description: (EC 2.6.--); glutamyl-tRNA (Gln) amidotransferase [validated,.  
 MUID:98004482]  
 C;Superfamily: PET112 protein  
 C;Keywords: aminotransferase

Query Match 72.7%; Score 32; DB 2; Length 476;  
 Best Local Similarity 75.0%; Pred. No. 52;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
 | | | |  
 Db 137 YSLVDFNR 144

RESULT 9  
 T44293

glutamyl-tRNA (Gln) amidotransferase subunit B BH0667 [imported] - *Bacillus*  
*halodurans* (strain C-125)  
 C;Species: *Bacillus halodurans*  
 C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
 C;Accession: T44293; C83733  
 R;Takami, H.; Nakasone, K.; Ogasawara, N.; Hiramata, C.; Nakamura, Y.; Masui, N.;  
 Fuji, F.; Takaki, Y.; Inoue, A.; Horikoshi, K.  
 Extremophiles 3, 29-34, 1999  
 A;Title: Sequencing of three lambda clones from the genome of alkaliphilic  
*Bacillus* sp. strain C-125.  
 A;Reference number: Z22745; MUID:99184646; PMID:10086842  
 A;Accession: T44293  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-476 <TAK>  
 A;Cross-references: UNIPROT:Q9Z9X0; EMBL:AB011836; NID:g4512345;  
 PIDN:BAA75312.1; PID:g4512347  
 A;Experimental source: strain C-125  
 R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji,  
 F.; Hiramata, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus*  
*halodurans* and genomic sequence comparison with *Bacillus subtilis*.

A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: C83733  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-476 <STO>  
A;Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BAB04386.1;  
GSPDB:GN0C137  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH0667  
A;Note: yerN  
C;Superfamily: PET112 protein

Query Match 72.7%; Score 32; DB 2; Length 476;  
Best Local Similarity 75.0%; Pred. No. 52;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| | | |  
Db 137 YSLVDFNR 144

#### RESULT 10

D71860

probable outer membrane protein - *Helicobacter pylori* (strain J99)

C;Species: *Helicobacter pylori*

A;Variety: strain J99

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004

C;Accession: D71860

R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.;  
Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge, B.L.; Carmel, G.; Tummino, P.J.;  
Caruso, A.; Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.; Merberg, D.;  
Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Trust, T.J.  
Nature 397, 176-180, 1999

A;Title: Genomic sequence comparison of two unrelated isolates of the human  
gastric pathogen *Helicobacter pylori*.

A;Reference number: A71800; MUID:99120557; PMID:9923682

A;Accession: D71860

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-751 <ARN>

A;Cross-references: UNIPROT:Q9ZKD1; GB:AE001529; GB:AE001439; NID:g4155590;  
PIDN:AAD06586.1; PID:g4155594

A;Experimental source: strain J99

C;Genetics:

A;Gene: jhp1008

Query Match 72.7%; Score 32; DB 2; Length 751;  
Best Local Similarity 57.1%; Pred. No. 84;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YGLVQFN 8  
| | : : | |  
Db 615 YGIIQYN 621

#### RESULT 11

E64842

probable monooxygenase b1007 - *Escherichia coli* (strain K-12)

C;Species: *Escherichia coli*

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C;Accession: E64842

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of *Escherichia coli* K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: E64842

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-152 <BLAT>

A;Cross-references: GB:AE000202; GB:U00096; NID:g1787233; PIDN:AAC74092.1; PID:g1787242; UWGP:b1007

A;Experimental source: strain K-12, substrain MG1655

C;Superfamily: 4-hydroxyphenylacetate 3-monooxygenase small chain

Query Match 70.5%; Score 31; DB 2; Length 152;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
||| |:  
Db 135 YGLVWFDR 142

RESULT 12

E90785

probable 4-hydroxyphenylacetate 3-monooxygenase ECs1253 [similarity] -

*Escherichia coli* (strain O157:H7, substrain RIMD 0509952)

C;Species: *Escherichia coli*

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 27-Nov-2001

C;Accession: E90785

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: E90785

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-152 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA034676.1; PID:g13360713; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs1253

C;Superfamily: 4-hydroxyphenylacetate 3-monooxygenase small chain

Query Match 70.5%; Score 31; DB 2; Length 152;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



Qy            2 YGLVQFNR 9  
             |||| |:  
Db           135 YGLVWFDR 142

RESULT 13

C85645

probable 4-hydroxyphenylacetate 3-monooxygenase Z1506 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)

C;Species: *Escherichia coli*

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C;Accession: C85645

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: C85645

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-152 <STO>

A;Cross-references: GB:AE005174; NID:g12514364; PIDN:AAG55623.1; GSPDB:GN00145; UWGP:Z1506

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: Z1506

C;Superfamily: 4-hydroxyphenylacetate 3-monooxygenase small chain

Query Match                    70.5%;   Score 31;   DB 2;   Length 152;  
Best Local Similarity       75.0%;   Pred. No. 26;  
Matches       6;   Conservative       1;   Mismatches       1;   Indels       0;   Gaps       0;

Qy            2 YGLVQFNR 9  
             |||| |:  
Db           135 YGLVWFDR 142

RESULT 14

S72490

N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) - *Bacillus stearothermophilus*

C;Species: *Bacillus stearothermophilus*

C;Date: 24-Oct-1998 #sequence\_revision 24-Oct-1998 #text\_change 09-Jul-2004

C;Accession: S72490; I39765

R;Savchenko, A.; Charlier, D.; Dion, M.; Weigel, P.; Hallet, J.N.; Holtham, C.; Baumberg, S.; Glansdorff, N.; Sakanyan, V.

Mol. Gen. Genet. 252, 69-78, 1996

A;Title: The arginine operon of *Bacillus stearothermophilus*: characterization of the control region and its interaction with the heterologous *B. subtilis* arginine repressor.

A;Reference number: S72490; MUID:96397511; PMID:8804405

A;Accession: S72490

A;Status: not compared with conceptual translation

A;Molecule type: DNA  
 A;Residues: 1-84 <SAV>  
 A;Cross-references: UNIPROT:Q07906  
 A;Experimental source: strain NCIB8224  
 R;Sakanyan, V.; Charlier, D.; Legrain, C.; Kochikyan, A.; Mett, I.; Pierard, P.; Glansdorff, N.  
 J. Gen. Microbiol. 139, 393-402, 1993  
 A;Title: Primary structure, partial purification and regulation of key enzymes of the acetyl cycle of arginine biosynthesis in *Bacillus stearothermophilus*: Dual function of ornithine acetyltransferase.  
 A;Reference number: I39765; MUID:93232760; PMID:8473852  
 A;Accession: I39765  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 52-345 <RES>  
 A;Cross-references: GB:L06036; NID:g304133; PIDN:AAA22196.1; PID:g304134  
 A;Experimental source: strain NCIB8224  
 C;Genetics:  
 A;Gene: argC  
 C;Superfamily: N-acetyl-gamma-glutamyl-phosphate reductase  
 C;Keywords: arginine biosynthesis; oxidoreductase  
 F;149/Active site: Cys #status predicted

Query Match 70.5%; Score 31; DB 2; Length 345;  
 Best Local Similarity 62.5%; Pred. No. 62;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 ||| ::||  
 Db 123 YGLTEWNR 135

#### RESULT 15

A70474

conserved hypothetical protein aq\_2027 - *Aquifex aeolicus*

C;Species: *Aquifex aeolicus*

C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004

C;Accession: A70474

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.  
 Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: A70474

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-364 <AQF>

A;Cross-references: UNIPROT:O67821; GB:AE000768; NID:g2984249; PIDN:AAC07788.1; PID:g2984261; GB:AE000657

A;Experimental source: strain VF5

C;Genetics:

A;Gene: aq\_2027

Query Match 70.5%; Score 31; DB 2; Length 364;  
 Best Local Similarity 62.5%; Pred. No. 65;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| |::|||  
Db 241 YNLLEFNR 248

Search completed: February 10, 2005, 15:59:23  
Job time : 14.9437 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 65.662 Seconds  
(without alignments)  
70.188 Million cell updates/sec

Title: US-10-067-484-3  
Perfect score: 44  
Sequence: 1 XYGLVQFNR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	38	86.4	678	2	Q64ZJ4	Q64zj4 bacteroides	
2	34	77.3	304	2	Q83AP7	Q83ap7 coxiella bu	
3	34	77.3	354	1	ARGC_BORBR	Q7wfc5 bordetella	
4	34	77.3	354	1	ARGC_BORPA	Q7w3z3 bordetella	
5	34	77.3	354	1	ARGC_BORPE	Q7vuw0 bordetella	
6	33	75.0	63	2	O27885	O27885 methanobact	

7	33	75.0	124	2	Q6Y228	Q6y228	pagrus majo
8	33	75.0	192	2	Q8TVJ3	Q8tvj3	methanopyru
9	33	75.0	253	1	RECO_STRA3	Q8e7x6	streptococc
10	33	75.0	253	1	RECO_STRA5	Q8e2g8	streptococc
11	33	75.0	313	2	O44967	O44967	caenorhabdi
12	33	75.0	333	2	Q8SRL7	Q8srl7	encephalito
13	33	75.0	533	2	Q8DHX4	Q8dhx4	synechococc
14	33	75.0	618	1	XYA2_BACST	P45704	bacillus st
15	33	75.0	662	2	Q9M2V6	Q9m2v6	arabidopsis
16	32	72.7	124	1	MKI1_HUMAN	Q9uha4	homo sapien
17	32	72.7	124	1	MKI1_MOUSE	O88653	mus musculu
18	32	72.7	291	2	Q9RDY3	Q9rdy3	legionella
19	32	72.7	293	2	Q82WJ7	Q82wj7	nitrosomona
20	32	72.7	296	1	YDEH_ECOLI	P31129	escherichia
21	32	72.7	296	2	Q8XB26	Q8xb26	escherichia
22	32	72.7	298	2	Q8FHD4	Q8fhd4	escherichia
23	32	72.7	299	2	Q7M8N4	Q7m8n4	wolinella s
24	32	72.7	321	2	Q6ASC1	Q6asc1	desulfotale
25	32	72.7	399	2	Q7PUT5	Q7put5	anopheles g
26	32	72.7	426	2	Q9LIY4	Q9liy4	oryza sativ
27	32	72.7	444	2	Q6Z9W0	Q6z9w0	oryza sativ
28	32	72.7	444	2	Q82BJ2	Q82bj2	streptomyce
29	32	72.7	475	2	Q9P8U4	Q9p8u4	aspergillus
30	32	72.7	476	1	GATB_BACHD	Q9z9x0	bacillus ha
31	32	72.7	476	1	GATB_BACST	Q931e1	bacillus st
32	32	72.7	476	1	GATB_BACSU	O30509	bacillus su
33	32	72.7	490	2	Q67KJ3	Q67kj3	symbiobacte
34	32	72.7	692	2	Q7NSR3	Q7nsr3	chromobacte
35	32	72.7	717	2	Q985I6	Q985i6	rhizobium l
36	32	72.7	733	2	Q872B5	Q872b5	neurospora
37	32	72.7	751	2	Q9ZKD1	Q9zkd1	helicobacte
38	32	72.7	761	2	Q89QM2	Q89qm2	bradyrhizob
39	32	72.7	810	2	Q7Q9F7	Q7q9f7	anopheles g
40	32	72.7	812	2	Q7RJ31	Q7rj31	plasmodium
41	32	72.7	949	2	Q8TZ35	Q8tz35	methanopyru
42	32	72.7	1026	2	Q7RSV2	Q7rsv2	plasmodium
43	32	72.7	1136	2	Q7RHC8	Q7rhc8	plasmodium
44	32	72.7	3228	2	Q6D920	Q6d920	erwinia car
45	31	70.5	57	2	Q8E4M8	Q8e4m8	streptococc

#### ALIGNMENTS

#### RESULT 1

Q64ZJ4

ID Q64ZJ4 PRELIMINARY; PRT; 678 AA.  
AC Q64ZJ4;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=BF0333;  
OS Bacteroides fragilis.  
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
OC Bacteroidaceae; Bacteroides.  
OX NCBI\_TaxID=817;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YCH46;  
 RA Kuwahara T., Yamashita A., Hirakawa H., Nakayama H., Toh H., Okada N.,  
 RA Kuhara S., Hattori M., Hayashi T., Ohnishi Y.;  
 RT "Genomic analysis of Bacteroides fragilis reveals extensive DNA  
 RT inversions regulating cell surface adaptation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2004).  
 DR EMBL; AP006841; BAD47082.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 678 AA; 77883 MW; DFFCC38F1938AAB0 CRC64;  
  
 Query Match 86.4%; Score 38; DB 2; Length 678;  
 Best Local Similarity 87.5%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
 Qy 2 YGLVQFNR 9  
 |||| |||  
 Db 583 YGLVDFNR 590

# RESULT 2

Q83AP7  
 ID Q83AP7 PRELIMINARY; PRT; 304 AA.  
 AC Q83AP7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Glucose-1-phosphate thymidyltransferase.  
 GN Name=rmlA; OrderedLocusNames=CBU1834;  
 OS Coxiella burnetii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;  
 OC Coxiellaceae; Coxiella.  
 OX NCBI\_TaxID=777;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Nine Mile phase I / RSA 493;  
 RX MEDLINE=22608657; PubMed=12704232; DOI=10.1073/pnas.0931379100;  
 RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,  
 RA Nelson W.C., Ward N.L., Tettelin H., Davidsen T.M., Beanan M.J.,  
 RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,  
 RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,  
 RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;  
 RT "Complete genome sequence of the Q-fever pathogen, Coxiella  
 RT burnetii.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).  
 CC -!- CATALYTIC ACTIVITY: dTTP + alpha-D-glucose 1-phosphate =  
 CC diphosphate + dTDP-glucose.  
 CC -!- SIMILARITY: Belongs to the glucose-1-phosphate  
 CC thymidyltransferase family.  
 DR EMBL; AE016965; AAO91327.1; -.  
 DR HSSP; P37744; 1H5R.  
 DR TIGR; CBU1834; -.  
 DR GO; GO:0008879; F:glucose-1-phosphate thymidyltransferase a. . .; IEA.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009058; P:biosynthesis; IEA.

DR GO; GO:0045226; P:extracellular polysaccharide biosynthesis; IEA.  
 DR InterPro; IPR005907; G1P\_thy\_trans\_1.  
 DR InterPro; IPR005835; NTP\_transferase.  
 DR Pfam; PF00483; NTP\_transferase; 1.  
 DR TIGRFAMs; TIGR01207; rmlA; 1.  
 KW Complete proteome; Kinase; Nucleotidyltransferase; Transferase.  
 SQ SEQUENCE 304 AA; 34288 MW; 60924BD5F5D4BFD0 CRC64;

Query Match 77.3%; Score 34; DB 2; Length 304;  
 Best Local Similarity 62.5%; Pred. No. 60;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 ||:||:  
 Db 143 YGVVEFNK 150

### RESULT 3

#### ARGC\_BORBR

ID ARGC\_BORBR STANDARD; PRT; 354 AA.  
 AC Q7WFC5;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-  
 DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).  
 GN Name=argC; OrderedLocusNames=BB4355;  
 OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=518;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RB50 / ATCC BAA-588;  
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;  
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
 RA Rabbinoiwitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 CC -!- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)  
 CC + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.  
 CC -!- PATHWAY: Arginine biosynthesis; third step.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily  
 CC 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; BX640450; CAE34718.1; -.  
DR HAMAP; MF\_00150; -; 1.  
DR InterPro; IPR000706; AGPR\_act\_site.  
DR InterPro; IPR011137; NAGSA\_deh.  
DR InterPro; IPR000534; Semialdh\_dh.  
DR Pfam; PF01118; Semialdehyde\_dh; 1.  
DR Pfam; PF02774; Semialdehyde\_dhC; 1.  
DR PIRSF; PIRSF000150; NAGSA\_deh; 1.  
DR ProDom; PD003765; AGPR\_act\_site; 1.  
DR TIGRFAMs; TIGR01850; argC; 1.  
DR PROSITE; PS01224; ARGC; FALSE\_NEG.  
KW Arginine biosynthesis; Complete proteome; NADP; Oxidoreductase.  
FT ACT\_SITE 156 156 By similarity.  
SQ SEQUENCE 354 AA; 37771 MW; 4E3CF0A5ABD2C7D4 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 354;  
Best Local Similarity 75.0%; Pred. No. 70;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
|||: ||  
Db 135 YGLVELNR 142

#### RESULT 4

##### ARGC\_BORPA

ID ARGC\_BORPA STANDARD; PRT; 354 AA.  
AC Q7W3Z3;  
DT 29-MAR-2004 (Rel. 43, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-  
DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).  
GN Name=argC; OrderedLocusNames=BPP3882;  
OS Bordetella parapertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=519;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=12822 / ATCC BAA-587;  
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;  
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,  
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,  
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,

RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
 RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
 RT Bordetella parapertussis and Bordetella bronchiseptica.";  
 RL Nat. Genet. 35:32-40(2003).  
 CC -!- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)  
 CC + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.  
 CC -!- PATHWAY: Arginine biosynthesis; third step.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -!- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily  
 CC 1.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; BX640435; CAE39165.1; -.  
 DR HAMAP; MF\_00150; -; 1.  
 DR InterPro; IPR000706; AGPR\_act\_site.  
 DR InterPro; IPR011137; NAGSA\_deh.  
 DR InterPro; IPR000534; Semialdh\_dh.  
 DR Pfam; PF01118; Semialdehyde\_dh; 1.  
 DR Pfam; PF02774; Semialdehyde\_dhC; 1.  
 DR PIRSF; PIRSF000150; NAGSA\_deh; 1.  
 DR ProDom; PD003765; AGPR\_act\_site; 1.  
 DR TIGRFAMs; TIGR01850; argC; 1.  
 DR PROSITE; PS01224; ARGC; FALSE\_NEG.  
 KW Arginine biosynthesis; Complete proteome; NADP; Oxidoreductase.  
 FT ACT\_SITE 156 156 By similarity.  
 SQ SEQUENCE 354 AA; 37771 MW; 4E3CF0A5ABD2C7D4 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 354;  
 Best Local Similarity 75.0%; Pred. No. 70;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 ||||: ||  
 Db 135 YGLVELNR 142

# RESULT 5

## ARGC\_BORPE

ID ARGC\_BORPE STANDARD; PRT; 354 AA.  
 AC Q7VUW0;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-  
 DE acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).  
 GN Name=argC; OrderedLocusNames=BP2960;  
 OS Bordetella pertussis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=520;



```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX      MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA      Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA      Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA      Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA      Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA      Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA      Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA      Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA      Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA      Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA      Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT      "Comparative analysis of the genome sequences of Bordetella pertussis,
RT      Bordetella parapertussis and Bordetella bronchiseptica.";
RL      Nat. Genet. 35:32-40(2003).
CC      -!- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+)
CC      + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.
CC      -!- PATHWAY: Arginine biosynthesis; third step.
CC      -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC      -!- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily
CC      1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; BX640420; CAE43232.1; -.
DR      HAMAP; MF_00150; -; 1.
DR      InterPro; IPR000706; AGPR_act_site.
DR      InterPro; IPR011137; NAGSA_deh.
DR      InterPro; IPR000534; Semialdh_dh.
DR      Pfam; PF01118; Semialdehyde_dh; 1.
DR      Pfam; PF02774; Semialdehyde_dhC; 1.
DR      PIRSF; PIRSF000150; NAGSA_deh; 1.
DR      ProDom; PD003765; AGPR_act_site; 1.
DR      TIGRFAMs; TIGR01850; argC; 1.
DR      PROSITE; PS01224; ARGC; FALSE_NEG.
KW      Arginine biosynthesis; Complete proteome; NADP; Oxidoreductase.
FT      ACT_SITE    156      156      By similarity.
SQ      SEQUENCE    354 AA;  37829 MW;  FB91310A11C6C402 CRC64;

```

```

Query Match          77.3%;  Score 34;  DB 1;  Length 354;
Best Local Similarity 75.0%;  Pred. No. 70;
Matches      6;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

Qy      2 YGLVQFNR 9
        ||||: ||
Db      135 YGLVELNR 142

```

RESULT 6

O27885

ID O27885 PRELIMINARY; PRT; 63 AA.  
AC O27885;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein MTH1857.  
GN OrderedLocusNames=MTH1857;  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanothermobacter.  
OX NCBI\_TaxID=187420;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Delta H;  
RX MEDLINE=98037514; PubMed=9371463;  
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,  
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,  
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;  
RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
RT deltaH: functional analysis and comparative genomics.";  
RL J. Bacteriol. 179:7135-7155(1997).  
DR EMBL; AE000938; AAB86323.1; -.  
DR PIR; D69115; D69115.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 63 AA; 7489 MW; DF86A1C75621D477 CRC64;

Query Match 75.0%; Score 33; DB 2; Length 63;  
Best Local Similarity 75.0%; Pred. No. 21;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFN 9  
||| :||  
Db 27 YGLVNWNR 34

RESULT 7

Q6Y228

ID Q6Y228 PRELIMINARY; PRT; 124 AA.  
AC Q6Y228;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Mitogen-activated protein kinase 1 interacting protein 1.  
OS Pagrus major (Red sea bream) (Chrysophrys major).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
OC Sparidae; Pagrus.  
OX NCBI\_TaxID=143350;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Chen S.L., Xu M.Y., Hu S.L., Li L.;

RT "Analysis of immune-relevant genes expressed in red sea bream  
RT spleen.";  
RL Aquaculture 240:115-130(2004).  
DR EMBL; AY190710; AAP20185.1; -.  
DR GO; GO:0016301; F:kinase activity; IEA.  
KW Kinase.  
SQ SEQUENCE 124 AA; 13721 MW; D6803482D85AAB18 CRC64;

Query Match 75.0%; Score 33; DB 2; Length 124;  
Best Local Similarity 75.0%; Pred. No. 41;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
| :|||||  
Db 78 YQIVQFNR 85

#### RESULT 8

Q8TVJ3

ID Q8TVJ3 PRELIMINARY; PRT; 192 AA.  
AC Q8TVJ3;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Uncharacterized protein.  
GN OrderedLocusNames=MK1396;  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
OC Methanopyrus.  
OX NCBI\_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014; DOI=10.1073/pnas.032671499;  
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
RT and monophyly of archaeal methanogens.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
DR EMBL; AE010432; AAM02609.1; -.  
KW Complete proteome.  
SQ SEQUENCE 192 AA; 20800 MW; 3542FB28505F4D5F CRC64;

Query Match 75.0%; Score 33; DB 2; Length 192;  
Best Local Similarity 75.0%; Pred. No. 63;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
|||:| |  
Db 134 YGLVKFER 141

#### RESULT 9

RECO\_STRA3

ID RECO\_STRA3 STANDARD; PRT; 253 AA.

AC Q8E7X6;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE DNA repair protein recO (Recombination protein O).  
 GN Name=recO; OrderedLocusNames=gbS0019;  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEM316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,  
 RA Kunst F.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513(2002).  
 CC -!- FUNCTION: Involved in DNA repair and recF pathway recombination  
 CC (By similarity).  
 CC -!- SIMILARITY: Belongs to the recO family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AL766843; CAD45664.1; -.  
 DR SagaList; gbs0019; -.  
 DR HAMAP; MF\_00201; -; 1.  
 DR InterPro; IPR003717; RecO.  
 DR Pfam; PF02565; RecO; 1.  
 KW Complete proteome; DNA recombination; DNA repair.  
 SQ SEQUENCE 253 AA; 29684 MW; 16C6FE56E86041A3 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 253;  
 Best Local Similarity 75.0%; Pred. No. 83;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 |||| :||  
 Db 7 YGLVLYNR 14

RESULT 10  
 RECO\_STRAS  
 ID RECO\_STRAS STANDARD; PRT; 253 AA.  
 AC Q8E2G8;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE DNA repair protein recO (Recombination protein O).

GN Name=recO; OrderedLocusNames=SAG0020;  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;  
 RA Tettelin H., Massignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
 RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,  
 RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
 RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,  
 RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,  
 RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,  
 RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).  
 CC -!- FUNCTION: Involved in DNA repair and recF pathway recombination  
 CC (By similarity).  
 CC -!- SIMILARITY: Belongs to the recO family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE014192; AAM98928.1; -.  
 DR TIGR; SAG0020; -.  
 DR HAMAP; MF\_00201; -; 1.  
 DR InterPro; IPR003717; RecO.  
 DR Pfam; PF02565; RecO; 1.  
 KW Complete proteome; DNA recombination; DNA repair.  
 SQ SEQUENCE 253 AA; 29684 MW; 16C6FE56E86041A3 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 253;  
 Best Local Similarity 75.0%; Pred. No. 83;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 YGLVQFNR 9  
 |||| :||  
 Db 7 YGLVLYNR 14

# RESULT 11

O44967

ID O44967 PRELIMINARY; PRT; 313 AA.  
 AC O44967;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Hypothetical protein C42C1.10.  
 GN Name=C42C1.10; ORFNames=C42C1.10;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RG WormBase Consortium;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Murray J., Rohlfing T., Davidson S., Wilson R.;  
 RT "The sequence of C. elegans cosmid C42C1.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RG WormBase Consortium;  
 RL Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Belongs to the mitochondrial carrier family.  
 DR EMBL; AF043695; AAL02464.1; -.  
 DR WormBase; WBGene00016588; C42C1.10.  
 DR WormPep; C42C1.10; CE29222.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005743; C:mitochondrial inner membrane; IEA.  
 DR GO; GO:0005488; F:binding; IEA.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR002113; Aden\_trnslctor.  
 DR InterPro; IPR001993; Mitoch\_carrier.  
 DR InterPro; IPR002067; Mit\_carrier.  
 DR Pfam; PF00153; Mito\_carr; 3.  
 DR PRINTS; PR00927; ADPTRNSLCASE.  
 DR PRINTS; PR00926; MITOCARRIER.  
 DR PROSITE; PS50920; SOLCAR; 3.  
 KW Hypothetical protein; Transmembrane; Transport.  
 SQ SEQUENCE 313 AA; 34570 MW; 1672FF49C4920C95 CRC64;

Query Match 75.0%; Score 33; DB 2; Length 313;  
 Best Local Similarity 85.7%; Pred. No. 1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 YGLVQFN 8  
 |||||:  
 Db 91 YGLVQFS 97

RESULT 12

Q8SRL7

ID Q8SRL7 PRELIMINARY; PRT; 333 AA.  
 AC Q8SRL7;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE POLYADENYLATE-BINDING PROTEIN 1.  
 GN Name=ECU07\_0340;  
 OS Encephalitozoon cuniculi GB-M1.  
 OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.  
 OX NCBI\_TaxID=284813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806; DOI=10.1038/35106579;  
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi.";  
 RL Nature 414:450-453 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL590447; CAD25566.1; -.  
 DR HSSP; P33240; 1P1T.  
 DR InterPro; IPR001209; Ribosomal\_S14.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; RRM\_1; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR PROSITE; PS00527; RIBOSOMAL\_S14; UNKNOWN\_1.  
 DR PROSITE; PS50102; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; UNKNOWN\_1.  
 SQ SEQUENCE 333 AA; 38017 MW; 3448E126F2C32253 CRC64;

Query Match 75.0%; Score 33; DB 2; Length 333;

Best Local Similarity 75.0%; Pred. No. 1.1e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9

|| |||:

Db 184 YGYVQFSR 191

RESULT 13

Q8DHX4

ID Q8DHX4 PRELIMINARY; PRT; 533 AA.  
 AC Q8DHX4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE NADH dehydrogenase subunit 4.

GN Name=ndhD2; OrderedLocusNames=tlr1819;  
 OS Synechococcus elongatus (Thermosynechococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=32046;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BP-1;  
 RX MEDLINE=22225144; PubMed=12240834;  
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,  
 RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the thermophilic cyanobacterium  
 RT Thermosynechococcus elongatus BP-1."  
 RL DNA Res. 9:123-130(2002).  
 DR EMBL; AP005375; BAC09371.1; -.  
 DR GO; GO:0009523; C:photosystem II; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR InterPro; IPR008948; L-Aspartase-like.  
 DR InterPro; IPR003918; NADHub\_oxred4.  
 DR InterPro; IPR010227; NDH\_I\_M.  
 DR InterPro; IPR001750; Oxidored\_q1.  
 DR Pfam; PF00361; Oxidored\_q1; 1.  
 DR PRINTS; PR01437; NUOXDRDTASE4.  
 DR TIGRFAMs; TIGR01972; NDH\_I\_M; 1.  
 KW Complete proteome; NAD; NADP; Oxidoreductase; Plastoquinone; Quinone.  
 SQ SEQUENCE 533 AA; 58068 MW; 2B5C46BD3AA9E9B2 CRC64;

Query Match 75.0%; Score 33; DB 2; Length 533;  
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 YGLVQFN 8  
 |||::||  
 Db 259 YGLIRFN 265

#### RESULT 14

##### XYA2\_BACST

ID XYA2\_BACST STANDARD; PRT; 618 AA.  
 AC P45704;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Beta-xylosidase precursor (EC 3.2.1.37) (1,4-beta-D-xylan  
 DE xylohydrolase) (Xylan 1,4-beta-xylosidase).  
 GN Name=xylA;  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
 OX NCBI\_TaxID=1422;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=No. 236;  
 RA Oh H., Choi Y.;  
 RT "Sequence analysis of B-xylosidase gene from Bacillus



RT stearothermophilus.";  
 RL Korean J. Appl. Microbiol. Biotechnol. 22:134-142(1994).  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-xylans so as to  
 CC remove successive D-xylose residues from the non-reducing termini.  
 CC -!- PATHWAY: Xylan degradation.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 52 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U15984; AAA50863.1; -.  
 DR InterPro; IPR000852; Glyco\_hydro\_52.  
 DR Pfam; PF03512; Glyco\_hydro\_52; 1.  
 DR PRINTS; PR00845; GLHYDRLASE52.  
 KW Glycosidase; Hydrolase; Signal; Xylan degradation.  
 FT SIGNAL 1 ? Potential.  
 FT CHAIN ? 618 Beta-xylosidase.  
 SQ SEQUENCE 618 AA; 69627 MW; 7D93B25CC8D03B33 CRC64;

Query Match 75.0%; Score 33; DB 1; Length 618;  
 Best Local Similarity 75.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 YGLVQFNR 9  
 | | | | |  
 Db 291 YALAQFNR 298

# RESULT 15

Q9M2V6

ID Q9M2V6 PRELIMINARY; PRT; 662 AA.  
 AC Q9M2V6;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE ABC transporter-like protein.  
 GN Name=T15C9\_100;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Barges M., Collado M.C., Navarro P., Terol J., Perez-Alonso M.,  
 RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X., Quetier F.,  
 RA Salanoubat M.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;

RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family.  
 DR EMBL; AL132970; CAB82705.1; -.  
 DR PIR; T47649; T47649.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; UNKNOWN\_1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 662 AA; 74382 MW; A846B787D4B2866B CRC64;

Query Match 75.0%; Score 33; DB 2; Length 662;  
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 GLVQFNR 9  
 |||:|||  
 Db 300 GLVEFNR 306

Search completed: February 10, 2005, 15:57:25  
 Job time : 68.662 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 61.0282 Seconds  
(without alignments)  
44.362 Million cell updates/sec

Title: US-10-067-484-4  
Perfect score: 34  
Sequence: 1 FYXFSTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	32	94.1	7	5	ABB81971	Abb81971 30 kDa ra
2	30	88.2	74	4	ABG11991	Abg11991 Novel hum
3	30	88.2	218	8	ADL04980	Adl04980 M. catarr
4	29	85.3	85	6	ABP80721	Abp80721 N. gonorr
5	29	85.3	85	6	ABP80255	Abp80255 N. gonorr
6	29	85.3	329	7	ADC94153	Adc94153 E. faeciu
7	29	85.3	358	6	ABU34617	Abu34617 Protein e
8	29	85.3	359	2	AAW46752	Aaw46752 D-alanine
9	29	85.3	364	6	ABP57038	Abp57038 Mycobacte

10	29	85.3	369	6	ABU33822	Abu33822	Protein e
11	29	85.3	373	6	ABP57037	Abp57037	Mycobacte
12	29	85.3	373	6	ABP57039	Abp57039	Mycobacte
13	29	85.3	373	6	ABP57036	Abp57036	Mycobacte
14	29	85.3	373	6	ABM15931	Abm15931	Mycobacte
15	29	85.3	373	6	ABU36891	Abu36891	Protein e
16	29	85.3	384	6	ABU35985	Abu35985	Protein e
17	29	85.3	509	7	ADF05659	Adf05659	Bacterial
18	29	85.3	635	4	ABB62805	Abb62805	Drosophil
19	28	82.4	20	5	AAM48298	Aam48298	Human CD4
20	28	82.4	30	5	AAM48295	Aam48295	Human CD4
21	28	82.4	30	5	AAM48294	Aam48294	Human CD4
22	28	82.4	40	4	AAM86125	Aam86125	Human imm
23	28	82.4	73	4	AAM17792	Aam17792	Peptide #
24	28	82.4	73	4	ABB36817	Abb36817	Peptide #
25	28	82.4	73	4	AAM30302	Aam30302	Peptide #
26	28	82.4	73	4	ABB31596	Abb31596	Peptide #
27	28	82.4	73	4	ABB22140	Abb22140	Protein #
28	28	82.4	73	4	AAM69961	Aam69961	Human bon
29	28	82.4	73	4	AAM57558	Aam57558	Human bra
30	28	82.4	73	4	ABG51661	Abg51661	Human liv
31	28	82.4	73	4	AAM05439	Aam05439	Peptide #
32	28	82.4	73	5	ABG39594	Abg39594	Human pep
33	28	82.4	86	4	AAO12999	Aao12999	Human pol
34	28	82.4	86	4	AAU30900	Aau30900	Novel hum
35	28	82.4	111	4	AAB63655	Aab63655	Human gas
36	28	82.4	130	4	AAM81232	Aam81232	Human hae
37	28	82.4	167	5	AAM48312	Aam48312	Protein r
38	28	82.4	175	3	AAG55603	Aag55603	Arabidops
39	28	82.4	230	6	ABU19098	Abu19098	Protein e
40	28	82.4	261	7	ADI60182	Adi60182	Secreted
41	28	82.4	271	3	AAG55602	Aag55602	Arabidops
42	28	82.4	271	6	ABU04648	Abu04648	Human exp
43	28	82.4	271	6	ABU04606	Abu04606	Human exp
44	28	82.4	273	5	AAM48308	Aam48308	Protein R
45	28	82.4	293	6	ABU04611	Abu04611	Human exp

#### ALIGNMENTS

##### RESULT 1

AEB81971

ID ABB81971 standard; peptide; 7 AA.

XX

AC ABB81971;

XX

DT 25-NOV-2002 (first entry)

XX

DE 30 kDa ragweed pollen allergen tryptic peptide 4.

XX

KW Ragweed; pollen; allergen; Ambt 7; glycoprotein; antiallergic;  
KW immunotherapy; disulphide protein.

XX

OS Ambrosia elatior.

XX

FH Key Location/Qualifiers

FT Misc-difference 3  
 FT /label= Leu or Ile  
 XX  
 PN WO200263012-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 04-FEB-2002; 2002WO-US003346.  
 XX  
 PR 05-FEB-2001; 2001US-0266686P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Buchanan BB, Del Val G, Frick OL;  
 XX  
 DR WPI; 2002-657539/70.  
 XX  
 PT New ragweed pollen allergens, useful in allergy testing and immunotherapy  
 PT regimens, particularly for treating sensitivity to pollen or pollen  
 PT allergy (e.g. anaphylaxis, or symptoms of hives or asthma) in a mammal;  
 PT especially a human.  
 XX  
 PS Claim 1; Page 53; 70pp; English.  
 XX  
 CC The invention relates to an isolated pollen allergen purified from  
 CC ragweed pollen, substantially free of any other pollen proteins, or a  
 CC protein that is an antigenic fragment of a pollen allergen Ambt 7. The  
 CC allergen is characterized by the following physiochemical and biological  
 CC properties: (a) being contained in pollen extracts; (b) a glycoprotein;  
 CC (c) a sulphhydryl group containing protein; (d) a molecular weight of  
 CC about 30 kDa as determined by SDS-polyacrylamide gel electrophoresis; and  
 CC (e) possessing allergen activity. The pollen allergen, or antigenic  
 CC protein fragment of the pollen allergen Ambt 7, or composition is useful  
 CC for treating sensitivity to pollen or pollen allergy in a mammal. This  
 CC allergy includes anaphylaxis or atopy, which includes the symptoms of hay  
 CC fever, asthma or hives. The allergen is also useful in allergy testing  
 CC and immunotherapy regimens. Sequences ABB81968-978 represent tryptic  
 CC peptide fragments of the 30 kDa ragweed complete pollen extract  
 CC disulphide protein allergen  
 XX  
 SQ Sequence 7 AA;

Query Match 94.1%; Score 32; DB 5; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FYXFSTK 7  
 |||||  
 Db 1 FYXFSTK 7

RESULT 2  
 ABG11991

ID ABG11991 standard; protein; 74 AA.  
 XX  
 AC ABG11991;  
 XX

DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #11982.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS76178.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 42350; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 74 AA;

Query Match

88.2%; Score 30; DB 4; Length 74;

Best Local Similarity 71.4%; Pred. No. 48;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:|  
Db 68 FYSFTTK 74

RESULT 3

ADL04980

ID ADL04980 standard; protein; 218 AA.

XX

AC ADL04980;

XX

DT 06-MAY-2004 (first entry)

XX

DE M. catarrhalis protein #746.

XX

KW Moraxella catarrhalis; infection.

XX

OS Moraxella catarrhalis.

XX

PN US6673910-B1.

XX

PD 06-JAN-2004.

XX

PF 04-APR-2000; 2000US-00540236.

XX

PR 08-APR-1999; 99US-0128416P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Breton GL;

XX

DR WPI; 2004-178127/17.

DR N-PSDB; ADL03060.

XX

PT New nucleic acid encoding a Moraxella catarrhalis polypeptide, useful for  
PT preparing a composition for diagnosing, preventing or treating infection  
PT caused by Moraxella catarrhalis.

XX

PS Disclosure; SEQ ID NO 2666; 429pp; English.

XX

CC The invention relates to an isolated nucleic acid encoding an Moraxella  
CC catarrhalis polypeptide. The nucleic acid is useful for preparing a  
CC composition for diagnosing, preventing or treating infection caused by  
CC Moraxella catarrhalis. The present sequence represents the amino acid  
CC sequence of a M. catarrhalis protein.

XX

SQ Sequence 218 AA;

Query Match 88.2%; Score 30; DB 8; Length 218;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:|

Db 129 FYSFATK 135

RESULT 4

ABP80721

ID ABP80721 standard; protein; 85 AA.

XX

AC ABP80721;

XX

DT 07-MAR-2003 (first entry)

XX

DE N. gonorrhoeae amino acid sequence SEQ ID 7972.

XX

KW Antibacterial; infection; vaccine; gene therapy.

XX

OS Neisseria gonorrhoeae.

XX

PN WO200279243-A2.

XX

PD 10-OCT-2002.

XX

PF 12-FEB-2002; 2002WO-IB002069.

XX

PR 12-FEB-2001; 2001GB-00003424.

XX

PA (CHIR-) CHIRON SPA.

XX

PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX

DR WPI; 2003-058415/05.

DR N-PSDB; ABZ41691.

XX

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection.

XX

PS Disclosure; Page 770; 815pp; English.

XX

CC The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
CC molecules of the invention

XX

SQ Sequence 85 AA;

Query Match 85.3%; Score 29; DB 6; Length 85;  
Best Local Similarity 71.4%; Pred. No. 90;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7

|: |||

Db 70 FFSFSTK 76



RESULT 5

ABP80255

ID ABP80255 standard; protein; 85 AA.

XX

AC ABP80255;

XX

DT 07-MAR-2003 (first entry)

XX

DE N. gonorrhoeae amino acid sequence SEQ ID 7040.

XX

KW Antibacterial; infection; vaccine; gene therapy.

XX

OS Neisseria gonorrhoeae.

XX

PN WO200279243-A2.

XX

PD 10-OCT-2002.

XX

PF 12-FEB-2002; 2002WO-IB002069.

XX

PR 12-FEB-2001; 2001GB-00003424.

XX

PA (CHIR-) CHIRON SPA.

XX

PI Fontana MR, Pizza M, Massignani V, Monaci E;

XX

DR WPI; 2003-058415/05.

DR N-PSDB; ABZ41225.

XX

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a

PT medicament for treating or preventing N. gonorrhoeae infection.

XX

PS Disclosure; Page 699; 815pp; English.

XX

CC The present invention relates to proteins from Neisseria gonorrhoeae.

CC Also disclosed are the nucleic acid molecules encoding the proteins and

CC antibodies that specifically bind to the proteins. The composition

CC comprising the protein, nucleic acid or antibody is useful for the

CC manufacture of a medicament for treating or preventing N. gonorrhoeae

CC infection, this may be in the form of a vaccine or gene therapy.

CC Sequences given in records ABP76736-ABP81046 represent nucleic acid

CC molecules of the invention

XX

SQ Sequence 85 AA;

Query Match 85.3%; Score 29; DB 6; Length 85;

Best Local Similarity 71.4%; Pred. No. 90;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7

|: |||

Db 70 FFSFSTK 76

RESULT 6

ADC94153

ID ADC94153 standard; protein; 329 AA.

XX  
 AC ADC94153;  
 XX  
 DT 01-JAN-2004 (first entry)  
 XX  
 DE E. faecium protein sequence SEQ ID 3780.  
 XX  
 KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
 KW abdominal-pelvic infection.  
 XX  
 OS Enterococcus faecium.  
 XX  
 PN US6583275-B1.  
 XX  
 PD 24-JUN-2003.  
 XX  
 PF 30-JUN-1998; 98US-00107532.  
 XX  
 PR 02-JUL-1997; 97US-0051571P.  
 PR 14-MAY-1998; 98US-0085598P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI; 2003-799836/75.  
 DR N-PSDB; ADC90499.  
 XX  
 PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
 PT Enterococcus faecium polypeptide useful for detection, prevention and  
 PT treatment of a pathological condition resulting from a bacterial  
 PT infection.  
 XX  
 PS Example 1; SEQ ID NO 3780; 243pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid derived from  
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
 CC one of 10 fully defined sequences given in the (or comprising 40  
 CC sequential nucleotides chosen from any of the nucleic acids, its  
 CC complement or sequences hybridising to it). Also included are a  
 CC recombinant vector comprising the nucleic acid operably linked to  
 CC transcription regulatory element, a cell comprising the vector and a  
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
 CC The nucleic acids is useful for diagnosing pathological conditions  
 CC resulting from E. faecium bacterial infection (e.g. urinary tract  
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
 CC infection) and for screening drugs such as agonists and antagonists. The  
 CC nucleic acid is useful for recombinant production of Candida albicans -  
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
 CC and vaccines containing the nucleic acid are useful for preventing or  
 CC treating Enterococcus faecium infections. The present sequence represents  
 CC one if the disclosed E. faecium proteins.  
 XX  
 SQ Sequence 329 AA;

Query Match

85.3%; Score 29; DB 7; Length 329;

Best Local Similarity 71.4%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
:| ||||  
Db 137 YYIFSTK 143

RESULT 7

ABU34617

ID ABU34617 standard; protein; 358 AA.

XX

AC ABU34617;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #20144.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Mycobacterium bovis.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

DR N-PSDB; ACA38487.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 25; SEQ ID NO 62541; 1766pp; English.

XX

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 358 AA;

Query Match 85.3%; Score 29; DB 6; Length 358;  
Best Local Similarity 71.4%; Pred. No. 3.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:  
Db 255 FYDFATK 261

#### RESULT 8

AAW46752

ID AAW46752 standard; protein; 359 AA.

XX

AC AAW46752;

XX

DT 08-JUN-1998 (first entry)

XX

DE D-alanine-D-alanine ligase sequence of *Mycobacterium avium*.

XX

KW D-alanine-D-alanine ligase; bacterial growth; alanine ligase;

KW drug screening; identification; antimicrobial agent; infection; antibody.

XX

OS *Mycobacterium avium*.

XX

PN WO9748809-A1.

XX

PD 24-DEC-1997.

XX

PF 10-JUN-1997; 97WO-EP003010.

XX

PR 18-JUN-1996; 96EP-00810405.

XX

PA (NOVS ) NOVARTIS AG.  
 XX  
 PI Oreilly T, Littlewood-Evans AJ;  
 XX  
 DR WPI; 1998-063147/06.  
 DR N-PSDB; AAV16298.  
 XX  
 PT New Mycobacterium avium D-ala-D-ala ligase - used to develop products for  
 PT identifying anti-microbial agents and for diagnostic and therapeutic  
 PT applications.  
 XX  
 PS Claim 2; Page 43-45; 55pp; English.  
 XX  
 CC The present sequence represents a D-alanine-D-alanine ligase of  
 CC Mycobacterium avium. This enzyme is a ubiquitous enzyme which is  
 CC essential for bacterial growth. Genomic DNA was extracted from M. avium,  
 CC and amplified using primers that were designed based on amino acid  
 CC homology of existing alanine ligases from E. coli, S. typhimurium, E.  
 CC faecalis and E. gallinarium to obtain the cDNA sequence. The M. avium D-  
 CC alanine-D-alanine ligase can be used in drug screening protocols for the  
 CC identification of antimicrobial agents suitable for treating M. avium  
 CC infections. Antibodies specific for the ligase can be used for diagnostic  
 CC and therapeutic applications  
 XX  
 SQ Sequence 359 AA;

Query Match 85.3%; Score 29; DB 2; Length 359;  
 Best Local Similarity 71.4%; Pred. No. 3.7e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYXFSTK 7  
 |||:  
 Db 264 FYDFATK 270

RESULT 9  
 ABP57038  
 ID ABP57038 standard; protein; 364 AA.  
 XX  
 AC ABP57038;  
 XX  
 DT 10-APR-2003 (first entry)  
 XX  
 DE Mycobacterium avium D-Ala-D-Ala ligase enzyme SEQ ID NO:44.  
 XX  
 KW D-Ala-D-Ala ligase; enzyme; bacterial; structure-based drug design;  
 KW protein co-ordinate data; D-Ala-D-Ala ligase inhibitor; antibacterial.  
 XX  
 OS Mycobacterium avium.  
 XX  
 PN WO2003002063-A2.  
 XX  
 PD 09-JAN-2003.  
 XX  
 PF 28-JUN-2002; 2002WO-US020465.  
 XX  
 PR 28-JUN-2001; 2001US-0301676P.

XX  
PA (ESSE-) ESSENTIAL THERAPEUTICS INC.  
PA (PLIV ) PLIVA DD ZAGREB.  
XX  
PI Navia MA, Ala PJ, Griffith JP, Ali JA, Faerman CH, Moe ST;  
PI Magee AS, Connelly PR, Perola E;  
XX  
DR WPI; 2003-201458/19.  
XX  
PT Evaluating association potential of chemical entity to complex having  
PT binding pocket defined by structural coordinates, by employing  
PT computational unit for entity-pocket fitting operation and analyzing the  
PT results.  
XX  
PS Example 8; Fig 10; 115pp; English.  
XX  
CC The present invention describes a method (M1) of evaluating the potential  
CC of a chemical entity (CE) to associate with a molecule or molecular  
CC complex comprising a binding pocket (BP) defined by specific structural  
CC coordinates (SC) of D-Ala-D-Ala ligase (I) E. coli amino acids Lys144,  
CC Glu180, Lys181, Leu183, Glu187, Asp257 and Glu270, by employing a  
CC computational unit to perform a fitting operation between CE and BP  
CC defined by SC and analysing the results of the fitting operation to  
CC quantify the association between CE and BP. Also described is a method  
CC (M2) for identifying a potential inhibitor of (I). M1 is useful for  
CC evaluating the potential of a chemical entity to associate with a  
CC molecule or molecular complex comprising a binding pocket. M2 is useful  
CC for identifying a potential inhibitor of D-Ala-D-Ala ligase. The methods  
CC are useful in the identification of key interaction in the active site of  
CC the enzyme, as well as the design and optimisation of inhibitors. The  
CC methods are also useful in the drug discovery methods, particularly for  
CC discovering new drugs that inhibit D-Ala-D-Ala ligase, an essential  
CC enzyme in the formation of bacterial cell walls. The present sequence  
CC represents a D-Ala-D-Ala ligase amino acid sequence given in an example  
CC from the present invention  
XX  
SQ Sequence 364 AA;

Query Match 85.3%; Score 29; DB 6; Length 364;  
Best Local Similarity 71.4%; Pred. No. 3.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| ||:  
Db 262 FYDFATK 268

RESULT 10  
ABU33822  
ID ABU33822 standard; protein; 369 AA.  
XX  
AC ABU33822;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #19349.  
XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS *Mycobacterium avium*.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA37692.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 61746; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of

CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 369 AA;

Query Match 85.3%; Score 29; DB 6; Length 369;  
Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:  
Db 266 FYDFATK 272

#### RESULT 11

ABP57037

ID ABP57037 standard; protein; 373 AA.

XX

AC ABP57037;

XX

DT 10-APR-2003 (first entry)

XX

DE Mycobacterium tuberculosis D-Ala-D-Ala ligase enzyme SEQ ID NO:43.

XX

KW D-Ala-D-Ala ligase; enzyme; bacterial; structure-based drug design;  
KW protein co-ordinate data; D-Ala-D-Ala ligase inhibitor; antibacterial.

XX

OS Mycobacterium tuberculosis.

XX

PN WO2003002063-A2.

XX

PD 09-JAN-2003.

XX

PF 28-JUN-2002; 2002WO-US020465.

XX

PR 28-JUN-2001; 2001US-0301676P.

XX

PA (ESSE-) ESSENTIAL THERAPEUTICS INC.

PA (PLIV ) PLIVA DD ZAGREB.

XX

PI Navia MA, Ala PJ, Griffith JP, Ali JA, Faerman CH, Moe ST;

PI Magee AS, Connelly PR, Perola E;

XX

DR WPI; 2003-201458/19.

XX

PT Evaluating association potential of chemical entity to complex having  
PT binding pocket defined by structural coordinates, by employing  
PT computational unit for entity-pocket fitting operation and analyzing the  
PT results.

XX

PS Example 8; Fig 10; 115pp; English.

XX

CC The present invention describes a method (M1) of evaluating the potential  
CC of a chemical entity (CE) to associate with a molecule or molecular  
CC complex comprising a binding pocket (BP) defined by specific structural



CC coordinates (SC) of D-Ala-D-Ala ligase (I) E. coli amino acids Lys144,  
 CC Glu180, Lys181, Leu183, Glu187, Asp257 and Glu270, by employing a  
 CC computational unit to perform a fitting operation between CE and BP  
 CC defined by SC and analysing the results of the fitting operation to  
 CC quantify the association between CE and BP. Also described is a method  
 CC (M2) for identifying a potential inhibitor of (I). M1 is useful for  
 CC evaluating the potential of a chemical entity to associate with a  
 CC molecule or molecular complex comprising a binding pocket. M2 is useful  
 CC for identifying a potential inhibitor of D-Ala-D-Ala ligase. The methods  
 CC are useful in the identification of key interaction in the active site of  
 CC the enzyme, as well as the design and optimisation of inhibitors. The  
 CC methods are also useful in the drug discovery methods, particularly for  
 CC discovering new drugs that inhibit D-Ala-D-Ala ligase, an essential  
 CC enzyme in the formation of bacterial cell walls. The present sequence  
 CC represents a D-Ala-D-Ala ligase amino acid sequence given in an example  
 CC from the present invention  
 XX  
 SQ Sequence 373 AA;

Query Match 85.3%; Score 29; DB 6; Length 373;  
 Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 || ||:  
 Db 270 FYDFATK 276

# RESULT 12

ABP57039

ID ABP57039 standard; protein; 373 AA.  
 XX  
 AC ABP57039;  
 XX  
 DT 10-APR-2003 (first entry)  
 XX  
 DE Mycobacterium smegmatis D-Ala-D-Ala ligase enzyme SEQ ID NO:45.  
 XX  
 KW D-Ala-D-Ala ligase; enzyme; bacterial; structure-based drug design;  
 KW protein co-ordinate data; D-Ala-D-Ala ligase inhibitor; antibacterial.  
 XX  
 OS Mycobacterium smegmatis.  
 XX  
 PN WO2003002063-A2.  
 XX  
 PD 09-JAN-2003.  
 XX  
 PF 28-JUN-2002; 2002WO-US020465.  
 XX  
 PR 28-JUN-2001; 2001US-0301676P.  
 XX  
 PA (ESSE-) ESSENTIAL THERAPEUTICS INC.  
 PA (PLIV ) PLIVA DD ZAGREB.  
 XX  
 PI Navia MA, Ala PJ, Griffith JP, Ali JA, Faerman CH, Moe ST;  
 PI Magee AS, Connelly PR, Perola E;  
 XX

DR WPI; 2003-201458/19.

XX

PT Evaluating association potential of chemical entity to complex having  
PT binding pocket defined by structural coordinates, by employing  
PT computational unit for entity-pocket fitting operation and analyzing the  
PT results.

XX

PS Example 8; Fig 10; 115pp; English.

XX

CC The present invention describes a method (M1) of evaluating the potential  
CC of a chemical entity (CE) to associate with a molecule or molecular  
CC complex comprising a binding pocket (BP) defined by specific structural  
CC coordinates (SC) of D-Ala-D-Ala ligase (I) E. coli amino acids Lys144,  
CC Glu180, Lys181, Leu183, Glu187, Asp257 and Glu270, by employing a  
CC computational unit to perform a fitting operation between CE and BP  
CC defined by SC and analysing the results of the fitting operation to  
CC quantify the association between CE and BP. Also described is a method  
CC (M2) for identifying a potential inhibitor of (I). M1 is useful for  
CC evaluating the potential of a chemical entity to associate with a  
CC molecule or molecular complex comprising a binding pocket. M2 is useful  
CC for identifying a potential inhibitor of D-Ala-D-Ala ligase. The methods  
CC are useful in the identification of key interaction in the active site of  
CC the enzyme, as well as the design and optimisation of inhibitors. The  
CC methods are also useful in the drug discovery methods, particularly for  
CC discovering new drugs that inhibit D-Ala-D-Ala ligase, an essential  
CC enzyme in the formation of bacterial cell walls. The present sequence  
CC represents a D-Ala-D-Ala ligase amino acid sequence given in an example  
CC from the present invention

XX

SQ Sequence 373 AA;

Query Match 85.3%; Score 29; DB 6; Length 373;  
Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYXFSTK 7  
|||:  
Db 270 FYDFATK 276

#### RESULT 13

ABP57036

ID ABP57036 standard; protein; 373 AA.

XX

AC ABP57036;

XX

DT 10-APR-2003 (first entry)

XX

DE Mycobacterium tuberculosis D-Ala-D-Ala ligase enzyme SEQ ID NO:42.

XX

KW D-Ala-D-Ala ligase; enzyme; bacterial; structure-based drug design;  
KW protein co-ordinate data; D-Ala-D-Ala ligase inhibitor; antibacterial.

XX

OS Mycobacterium tuberculosis.

XX

PN WO2003002063-A2.

XX

PD 09-JAN-2003.  
 XX  
 PF 28-JUN-2002; 2002WO-US020465.  
 XX  
 PR 28-JUN-2001; 2001US-0301676P.  
 XX  
 PA (ESSE-) ESSENTIAL THERAPEUTICS INC.  
 PA (PLIV ) PLIVA DD ZAGREB.  
 XX  
 PI Navia MA, Ala PJ, Griffith JP, Ali JA, Faerman CH, Moe ST;  
 PI Magee AS, Connelly PR, Perola E;  
 XX  
 DR WPI; 2003-201458/19.  
 XX  
 PT Evaluating association potential of chemical entity to complex having  
 PT binding pocket defined by structural coordinates, by employing  
 PT computational unit for entity-pocket fitting operation and analyzing the  
 PT results.  
 XX  
 PS Example 8; Fig 10; 115pp; English.  
 XX  
 CC The present invention describes a method (M1) of evaluating the potential  
 CC of a chemical entity (CE) to associate with a molecule or molecular  
 CC complex comprising a binding pocket (BP) defined by specific structural  
 CC coordinates (SC) of D-Ala-D-Ala ligase (I) E. coli amino acids Lys144,  
 CC Glu180, Lys181, Leu183, Glu187, Asp257 and Glu270, by employing a  
 CC computational unit to perform a fitting operation between CE and BP  
 CC defined by SC and analysing the results of the fitting operation to  
 CC quantify the association between CE and BP. Also described is a method  
 CC (M2) for identifying a potential inhibitor of (I). M1 is useful for  
 CC evaluating the potential of a chemical entity to associate with a  
 CC molecule or molecular complex comprising a binding pocket. M2 is useful  
 CC for identifying a potential inhibitor of D-Ala-D-Ala ligase. The methods  
 CC are useful in the identification of key interaction in the active site of  
 CC the enzyme, as well as the design and optimisation of inhibitors. The  
 CC methods are also useful in the drug discovery methods, particularly for  
 CC discovering new drugs that inhibit D-Ala-D-Ala ligase, an essential  
 CC enzyme in the formation of bacterial cell walls. The present sequence  
 CC represents a D-Ala-D-Ala ligase amino acid sequence given in an example  
 CC from the present invention  
 XX  
 SQ Sequence 373 AA;

Query Match 85.3%; Score 29; DB 6; Length 373;  
 Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYXFSTK 7  
 || ||:  
 Db 270 FYDFATK 276

#### RESULT 14

ABM15931

ID ABM15931 standard; protein; 373 AA.

XX

AC ABM15931;

XX  
 DT 26-SEP-2003 (first entry)  
 XX  
 DE Mycobacterium tuberculosis mycobacterial antigen protein SEQ ID NO:227.  
 XX  
 KW Mycobacterium tuberculosis; mycobacterial; antigen; infection; vaccine;  
 KW tuberculostatic; mycobacterial peptide; mycobacterial infection.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 PN WO2003033530-A2.  
 XX  
 PD 24-APR-2003.  
 XX  
 PF 14-OCT-2002; 2002WO-GB004647.  
 XX  
 PR 12-OCT-2001; 2001GB-00024593.  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI James B, Bacon J, March P;  
 XX  
 DR WPI; 2003-393501/37.  
 DR N-PSDB; ACF39425.  
 XX  
 PT New isolated mycobacterial peptide encoded by a gene that is induced or  
 PT up-regulated under high oxygen tension, useful for diagnosing, treating  
 PT or preventing a mycobacterial infection.  
 XX  
 PS Claim 1; Page 368-369; 392pp; English.  
 XX  
 CC The present invention describes an isolated mycobacterial peptide (I), or  
 CC its fragment, variant or derivative encoded by a gene whose expression is  
 CC induced or up-regulated during culture of a mycobacterium under  
 CC continuous culture conditions of a dissolved oxygen tension of at least  
 CC 30% air saturation measured at 37 plus degrees Celsius when compared with  
 CC a dissolved oxygen tension of up to 10% air saturation measured at 37  
 CC plus degrees Celsius. (I) has tuberculostatic activity and can be used in  
 CC vaccines. The mycobacterial peptide (I) or its fragment, variant or  
 CC derivative, inhibitor, antibody, attenuated mycobacterium, attenuated  
 CC microbial carrier, DNA sequence, DNA plasmid, RNA sequence, or RNA vector  
 CC from the present invention can be used for manufacturing a medicament for  
 CC treating or preventing a mycobacterial infection. The peptide or its  
 CC fragment, variant or derivative, the antibody, or a polynucleotide probe  
 CC comprising at least 8 nucleotides, where the probe binds to at least a  
 CC part of the gene, is useful for manufacturing a diagnostic reagent for  
 CC identifying a mycobacterial infection. The present sequence represents a  
 CC Mycobacterium tuberculosis mycobacterial antigen, which is used in the  
 CC exemplification of the present invention  
 XX  
 SQ Sequence 373 AA;

Query Match 85.3%; Score 29; DB 6; Length 373;  
 Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7

Db

|| |:||  
270 FYDFATK 276

RESULT 15

ABU36891

ID ABU36891 standard; protein; 373 AA.

XX

AC ABU36891;

XX

DT 19-JUN-2003 (first entry)

XX

DE Protein encoded by Prokaryotic essential gene #22418.

XX

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX

OS Mycobacterium tuberculosis.

XX

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WG-US009107.

XX

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX

PA (ELIT-) ELITRA PHARM INC.

XX

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX

DR WPI; 2003-029926/02.

DR N-PSDB; ACA40761.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

XX

PS Claim 25; SEQ ID NO 64815; 1766pp; English.

XX

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 373 AA;

Query Match 85.3%; Score 29; DB 6; Length 373;  
Best Local Similarity 71.4%; Pred. No. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:|  
Db 270 FYDFATK 276

Search completed: February 10, 2005, 15:48:41  
Job time : 63.0282 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 15.6761 Seconds  
(without alignments)  
33.334 Million cell updates/sec

Title: US-10-067-484-4  
Perfect score: 34  
Sequence: 1 FYXFSTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	% Query		DB	ID	Description
	Score	Match Length			
1	30	88.2	218	4	US-09-540-236-2666 Sequence 2666, Ap
2	29	85.3	63	4	US-09-248-796A-24322 Sequence 24322, A
3	29	85.3	329	4	US-09-107-532A-3780 Sequence 3780, Ap
4	29	85.3	509	4	US-09-543-681A-5944 Sequence 5944, Ap
5	28	82.4	177	3	US-08-938-669A-30 Sequence 30, Appl
6	28	82.4	177	4	US-09-306-828-30 Sequence 30, Appl
7	28	82.4	295	4	US-09-949-016-9143 Sequence 9143, Ap
8	28	82.4	361	1	US-07-946-497-6 Sequence 6, Appli
9	28	82.4	361	1	US-08-483-322-6 Sequence 6, Appli
10	28	82.4	361	2	US-08-478-882-6 Sequence 6, Appli
11	28	82.4	361	4	US-09-949-016-5968 Sequence 5968, Ap
12	28	82.4	464	2	US-09-021-323-3 Sequence 3, Appli
13	28	82.4	573	4	US-09-248-796A-15753 Sequence 15753, A
14	27	79.4	75	4	US-09-248-796A-17035 Sequence 17035, A
15	27	79.4	85	4	US-09-543-681A-5659 Sequence 5659, Ap
16	27	79.4	259	4	US-09-583-110-2888 Sequence 2888, Ap
17	27	79.4	285	4	US-09-107-433-4723 Sequence 4723, Ap
18	27	79.4	356	4	US-09-270-767-37087 Sequence 37087, A
19	27	79.4	356	4	US-09-270-767-52304 Sequence 52304, A
20	27	79.4	367	4	US-09-543-681A-4571 Sequence 4571, Ap
21	27	79.4	487	4	US-09-620-412C-349 Sequence 349, App
22	27	79.4	487	4	US-09-598-419-349 Sequence 349, App
23	27	79.4	501	4	US-09-489-039A-13806 Sequence 13806, A
24	27	79.4	585	4	US-09-328-352-6426 Sequence 6426, Ap
25	27	79.4	821	4	US-09-556-877-195 Sequence 195, App
26	27	79.4	821	4	US-09-620-412C-195 Sequence 195, App
27	27	79.4	821	4	US-09-598-419-195 Sequence 195, App
28	27	79.4	1776	4	US-09-556-877-179 Sequence 179, App
29	27	79.4	1776	4	US-09-620-412C-179 Sequence 179, App
30	27	79.4	1776	4	US-09-598-419-179 Sequence 179, App
31	26	76.5	213	4	US-09-252-991A-23391 Sequence 23391, A
32	26	76.5	263	4	US-09-252-991A-24015 Sequence 24015, A
33	26	76.5	268	4	US-09-248-796A-19955 Sequence 19955, A
34	26	76.5	291	4	US-10-083-624-2 Sequence 2, Appli
35	26	76.5	310	4	US-09-543-681A-6858 Sequence 6858, Ap
36	26	76.5	324	4	US-09-252-991A-31493 Sequence 31493, A

37	26	76.5	331	4	US-09-248-796A-18164	Sequence 18164, A
38	26	76.5	364	1	US-08-454-196-9	Sequence 9, Appli
39	26	76.5	364	1	US-08-454-196-10	Sequence 10, Appl
40	26	76.5	364	2	US-08-286-819A-32	Sequence 32, Appl
41	26	76.5	364	3	US-08-980-357-32	Sequence 32, Appl
42	26	76.5	364	3	US-09-064-033-9	Sequence 9, Appli
43	26	76.5	364	3	US-09-064-033-10	Sequence 10, Appl
44	26	76.5	364	4	US-09-291-046-9	Sequence 9, Appli
45	26	76.5	364	4	US-09-291-046-10	Sequence 10, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-540-236-2666

```
; Sequence 2666, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
MORAXELLA CATARRHALIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2666
; LENGTH: 218
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2666
```

```
Query Match          88.2%;  Score 30;  DB 4;  Length 218;
Best Local Similarity 71.4%;  Pred. No. 23;
Matches      5;  Conservative      1;  Mismatches      1;  Indels      0;  Gaps      0;
```

```
Qy      1 FYXFSTK 7
        |||:|
Db      129 FYSFATK 135
```

##### RESULT 2

US-09-248-796A-24322

```
; Sequence 24322, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA
ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
```



; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 24322  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-24322

Query Match 85.3%; Score 29; DB 4; Length 63;  
Best Local Similarity 71.4%; Pred. No. 11;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| ||:  
Db 25 FYTFSSK 31

RESULT 3

US-09-107-532A-3780

; Sequence 3780, Application US/09107532A  
; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND  
THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 3780:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 329 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...329  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3780:  
US-09-107-532A-3780

Query Match 85.3%; Score 29; DB 4; Length 329;  
Best Local Similarity 71.4%; Pred. No. 60;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
: | |||  
Db 137 YYIFSTK 143

RESULT 4

US-09-543-681A-5944  
; Sequence 5944, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS  
MIRABILIS FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5944  
; LENGTH: 509  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5944

Query Match 85.3%; Score 29; DB 4; Length 509;  
Best Local Similarity 71.4%; Pred. No. 95;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| :|||  
Db 43 FYNYSTK 49

RESULT 5

US-08-938-669A-30  
; Sequence 30, Application US/08938669A  
; Patent No. 6171788  
; GENERAL INFORMATION:  
; APPLICANT: Nguyen, Thai D.

; APPLICANT: Polansky, Jon R.  
 ; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS,  
 ; TITLE OF INVENTION: PROGNOSIS AND TREATMENT OF GLAUCOMA AND  
 ; TITLE OF INVENTION: RELATED DISEASES  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Howrey & Simon  
 ; STREET: 1299 Pennsylvania Avenue, N.W.  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20004-2402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/938,669A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/791,154  
 ; FILING DATE: 28-JAN-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mendelson, Elliot  
 ; REGISTRATION NUMBER: P-42,878  
 ; REFERENCE/DOCKET NUMBER: 07425-0034  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202 383-6857  
 ; TELEFAX: 202 383-6610  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 30:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 177 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: No. 6171788e  
 US-08-938-669A-30

Query Match 82.4%; Score 28; DB 3; Length 177;  
 Best Local Similarity 71.4%; Pred. No. 54;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 || ||  
 Db 126 FYMFDTK 132

RESULT 6  
 US-09-306-828-30  
 ; Sequence 30, Application US/09306828  
 ; Patent No. 6475724  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nguyen, Thai D.  
 ; APPLICANT: Polansky, Jon R.

; APPLICANT: Chen, Pu  
; APPLICANT: Chen, Hua  
; TITLE OF INVENTION: Nucleic Acids, Kits, And Methods For The Diagnosis,  
Prognosis And Treatment Of Glaucoma And Related Dis; FILE REFERENCE:  
07425.0057.US01  
; CURRENT APPLICATION NUMBER: US/09/306,828  
; CURRENT FILING DATE: 1999-05-07  
; EARLIER APPLICATION NUMBER: US 09/227,881  
; EARLIER FILING DATE: 1999-01-11  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: Microsoft Word 97  
; SEQ ID NO 30  
; LENGTH: 177  
; TYPE: PRT  
; ORGANISM: Rana catesbeiana  
US-09-306-828-30

Query Match 82.4%; Score 28; DB 4; Length 177;  
Best Local Similarity 71.4%; Pred. No. 54;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||  
Db 126 FYMFDTK 132

#### RESULT 7

US-09-949-016-9143  
; Sequence 9143, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES  
THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 9143  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-9143

Query Match 82.4%; Score 28; DB 4; Length 295;  
Best Local Similarity 83.3%; Pred. No. 91;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFST 6

Db

|| |||  
239 FYTFST 244

RESULT 8

US-07-946-497-6

; Sequence 6, Application US/07946497

; Patent No. 5506119

; GENERAL INFORMATION:

; APPLICANT: HERRLICH, Peter

; APPLICANT: PONTA, Helmut

; APPLICANT: GUENTHERT, Ursula

; APPLICANT: MATZKU, Siegfried

; APPLICANT: WENZL, Achim

; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE  
PROTEINS,

; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/946,497

; FILING DATE: 19921109

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 16915/145

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: hCD44

US-07-946-497-6

Query Match 82.4%; Score 28; DB 1; Length 361;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy

1 FYXFST 6

|| |||

Db 195 FYTFST 200

RESULT 9

US-08-483-322-6

; Sequence 6, Application US/08483322

; Patent No. 5760178

; GENERAL INFORMATION:

; APPLICANT: HERRLICH, Peter

; APPLICANT: PONTA, Helmut

; APPLICANT: GUENTHERT, Ursula

; APPLICANT: MATZKU, Siegfried

; APPLICANT: WENZL, Achim

; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE  
PROTEINS,

; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/483,322

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/946,497

; FILING DATE: 09-NOV-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 16915/145

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: hCD44

US-08-483-322-6

Query Match 82.4%; Score 28; DB 1; Length 361;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy            1 FYXFST 6  
              || |||  
Db            195 FYTFST 200

RESULT 10

US-08-478-882-6

; Sequence 6, Application US/08478882

; Patent No. 5885575

; GENERAL INFORMATION:

; APPLICANT: HERRLICH, Peter

; APPLICANT: PONTA, Helmut

; APPLICANT: GUENTHERT, Ursula

; APPLICANT: MATZKU, Siegfried

; APPLICANT: WENZL, Achim

; TITLE OF INVENTION: VARIANT CD44 SURFACE PROTEINS, DNA

; TITLE OF INVENTION: SEQUENCES CODING THESE, ANTIBODIES AGAINST THESE  
PROTEINS,

; TITLE OF INVENTION: AS WELL AS THEIR USE IN DIAGNOSIS AND THERAPY

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington, D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/478,882

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/07/946,497

; FILING DATE: 19921109

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 16915/145

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 361 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; CLONE: hCD44

US-08-478-882-6

Query Match                    82.4%;   Score 28;   DB 2;   Length 361;  
Best Local Similarity        83.3%;   Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFST 6  
|| |||  
Db 195 FYTFST 200

RESULT 11

US-09-949-016-5968

; Sequence 5968, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 5968

; LENGTH: 361

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-5968

Query Match 82.4%; Score 28; DB 4; Length 361;

Best Local Similarity 83.3%; Pred. No. 1.1e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFST 6  
|| |||  
Db 195 FYTFST 200

RESULT 12

US-09-021-323-3

; Sequence 3, Application US/09021323

; Patent No. 5929033

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Corley, Neil C.

; APPLICANT: Yue, Henry

; TITLE OF INVENTION: EXTRACELLULAR MUCOUS MATRIX

; TITLE OF INVENTION: GLYCOPROTEIN

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto



; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/021,323  
 ; FILING DATE: Filed Herewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Billings, Lucy J.  
 ; REGISTRATION NUMBER: 36,749  
 ; REFERENCE/DOCKET NUMBER: PF-0477 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-855-0555  
 ; TELEFAX: 650-845-4166  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 464 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GenBank  
 ; CLONE: 294502  
 US-09-021-323-3

Query Match 82.4%; Score 28; DB 2; Length 464;  
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 || ||  
 Db 411 FYMFDTK 417

RESULT 13  
 US-C9-248-796A-15753  
 ; Sequence 15753, Application US/09248796A  
 ; Patent No. 6747137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Keith Weinstock et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA  
 ALBICANS  
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248,796A  
 ; CURRENT FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: US 60/074,725  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 15753  
; LENGTH: 573  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-15753

Query Match 82.4%; Score 28; DB 4; Length 573;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| |||:  
Db 114 FYKFSTE 120

RESULT 14

US-09-248-796A-17035  
; Sequence 17035, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17035  
; LENGTH: 75  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17035

Query Match 79.4%; Score 27; DB 4; Length 75;  
Best Local Similarity 83.3%; Pred. No. 38;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFST 6  
|| |||  
Db 1 FYQFST 6

RESULT 15

US-09-543-681A-5659  
; Sequence 5659, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5659  
; LENGTH: 85  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5659

Query Match 79.4%; Score 27; DB 4; Length 85;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||  
Db 20 FYYFPTK 26

Search completed: February 10, 2005, 16:02:08  
Job time : 16.6761 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:49:10 ; Search time 41.9014 Seconds  
(without alignments)  
54.586 Million cell updates/sec

Title: US-10-067-484-4  
Perfect score: 34  
Sequence: 1 FYXFSTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*  
 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	32	94.1	7	14	US-10-067-484-4	Sequence 4, Appli
2	32	94.1	7	14	US-10-067-620-4	Sequence 4, Appli
3	29	85.3	82	15	US-10-424-599-240774	Sequence 240774,
4	29	85.3	83	16	US-10-437-963-165869	Sequence 165869,
5	29	85.3	85	15	US-10-424-599-152381	Sequence 152381,
6	29	85.3	86	15	US-10-424-599-222727	Sequence 222727,
7	29	85.3	111	15	US-10-424-599-229590	Sequence 229590,
8	29	85.3	284	15	US-10-425-114-55509	Sequence 55509, A
9	29	85.3	358	15	US-10-282-122A-62541	Sequence 62541, A
10	29	85.3	364	14	US-10-186-886-44	Sequence 44, Appl
11	29	85.3	369	15	US-10-282-122A-61746	Sequence 61746, A
12	29	85.3	373	14	US-10-186-886-42	Sequence 42, Appl
13	29	85.3	373	14	US-10-186-886-43	Sequence 43, Appl
14	29	85.3	373	14	US-10-186-886-45	Sequence 45, Appl
15	29	85.3	373	15	US-10-282-122A-64815	Sequence 64815, A
16	29	85.3	384	15	US-10-282-122A-63909	Sequence 63909, A
17	29	85.3	558	15	US-10-424-599-256660	Sequence 256660,
18	28	82.4	27	15	US-10-425-114-45238	Sequence 45238, A
19	28	82.4	69	15	US-10-424-599-268087	Sequence 268087,
20	28	82.4	73	9	US-09-864-761-37438	Sequence 37438, A
21	28	82.4	73	15	US-10-424-599-247522	Sequence 247522,
22	28	82.4	88	15	US-10-424-599-175266	Sequence 175266,
23	28	82.4	97	15	US-10-424-599-174268	Sequence 174268,
24	28	82.4	106	15	US-10-424-599-194236	Sequence 194236,
25	28	82.4	121	15	US-10-424-599-252659	Sequence 252659,
26	28	82.4	130	9	US-09-796-692-1596	Sequence 1596, Ap
27	28	82.4	130	14	US-10-040-862-1596	Sequence 1596, Ap
28	28	82.4	130	15	US-10-057-475B-1596	Sequence 1596, Ap
29	28	82.4	130	15	US-10-154-884B-1596	Sequence 1596, Ap
30	28	82.4	130	16	US-10-764-324-1596	Sequence 1596, Ap

31	28	82.4	135	15	US-10-424-599-187725	Sequence 187725,
32	28	82.4	172	15	US-10-424-599-167569	Sequence 167569,
33	28	82.4	177	14	US-10-244-633-30	Sequence 30, Appl
34	28	82.4	189	16	US-10-767-701-34477	Sequence 34477, A
35	28	82.4	230	15	US-10-282-122A-47022	Sequence 47022, A
36	28	82.4	303	15	US-10-424-599-181047	Sequence 181047,
37	28	82.4	303	15	US-10-425-114-40861	Sequence 40861, A
38	28	82.4	334	15	US-10-424-599-247327	Sequence 247327,
39	28	82.4	359	9	US-09-864-921-99	Sequence 99, Appl
40	28	82.4	360	9	US-09-738-626-4958	Sequence 4958, Ap
41	28	82.4	361	11	US-09-836-544-32	Sequence 32, Appl
42	28	82.4	361	16	US-10-663-244-3	Sequence 3, Appli
43	28	82.4	361	16	US-10-741-601-340	Sequence 340, App
44	28	82.4	361	17	US-10-741-600-1048	Sequence 1048, Ap
45	28	82.4	365	15	US-10-282-122A-53732	Sequence 53732, A

#### ALIGNMENTS

##### RESULT 1

US-10-067-484-4  
 ; Sequence 4, Application US/10067484  
 ; Publication No. US20030170763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Buchanan, Bob B.  
 ; APPLICANT: del Val, Gregorio  
 ; APPLICANT: Frick, Oscar L.  
 ; TITLE OF INVENTION: RAGWEED ALLERGENS  
 ; FILE REFERENCE: 416272000200  
 ; CURRENT APPLICATION NUMBER: US/10/067,484  
 ; CURRENT FILING DATE: 2002-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/266,686  
 ; PRIOR FILING DATE: 2001-02-05  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 7  
 ; TYPE: PRT  
 ; ORGANISM: Ragweed  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: 3  
 ; OTHER INFORMATION: Xaa= Leucine or Isoleucine  
 US-10-067-484-4

Query Match 94.1%; Score 32; DB 14; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	FYXFSTK	7
Db	1	FYXFSTK	7

##### RESULT 2

US-10-067-620-4

```
; Sequence 4, Application US/10067620
; Publication No. US20030180225A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; APPLICANT: Teuber, Suzanne S.
; TITLE OF INVENTION: WALNUT AND RYEGRASS ALLERGENS
; FILE REFERENCE: 416272003400
; CURRENT APPLICATION NUMBER: US/10/067,620
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Ragweed
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 3
; OTHER INFORMATION: Xaa= Leucine or Isoleucine
US-10-067-620-4
```

```
Query Match          94.1%; Score 32; DB 14; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.2e+06;
Matches      7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;
```

```
QY      1 FYXFSTK 7
        |||||
Db      1 FYXFSTK 7
```

# RESULT 3

```
US-10-424-599-240774
; Sequence 240774, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 240774
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_59446C.1.pep
US-10-424-599-240774
```

Query Match 85.3%; Score 29; DB 15; Length 82;  
Best Local Similarity 71.4%; Pred. No. 1e+02;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| ||  
Db 24 FYTFKTK 30

RESULT 4

US-10-437-963-165869

; Sequence 165869, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules  
Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 165869

; LENGTH: 83

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_64633C.1.pep

US-10-437-963-165869

Query Match 85.3%; Score 29; DB 16; Length 83;  
Best Local Similarity 71.4%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| ||:  
Db 11 FYAFSSK 17

RESULT 5

US-10-424-599-152381

; Sequence 152381, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 152381  
 ; LENGTH: 85  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_108623C.1.pep  
 US-10-424-599-152381

Query Match 85.3%; Score 29; DB 15; Length 85;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 || ||:  
 Db 63 FYFFATK 69

# RESULT 6

US-10-424-599-222727  
 ; Sequence 222727, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 222727  
 ; LENGTH: 86  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_43150C.1.pep  
 US-10-424-599-222727

Query Match 85.3%; Score 29; DB 15; Length 86;  
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 || ||:  
 Db 64 FYSFSTQ 70



RESULT 7

US-10-424-599-229590

; Sequence 229590, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 229590

; LENGTH: 111

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_49344C.1.pep

US-10-424-599-229590

Query Match 85.3%; Score 29; DB 15; Length 111;

Best Local Similarity 71.4%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7

:| |||

Db 75 YYLFSTK 81

RESULT 8

US-10-425-114-55509

; Sequence 55509, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 55509

; LENGTH: 284

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY067A01\_FLI.pep

US-10-425-114-55509

Query Match 85.3%; Score 29; DB 15; Length 284;  
Best Local Similarity 71.4%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| |||:  
Db 228 FYSFSTE 234

RESULT 9

US-10-282-122A-62541

; Sequence 62541, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62541  
; LENGTH: 358

; TYPE: PRT  
; ORGANISM: Mycobacterium bovis  
US-10-282-122A-62541

Query Match 85.3%; Score 29; DB 15; Length 358;  
Best Local Similarity 71.4%; Pred. No. 4.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:|  
Db 255 FYDFATK 261

RESULT 10

US-10-186-886-44

; Sequence 44, Application US/10186886  
; Publication No. US20030119061A1  
; GENERAL INFORMATION:  
; APPLICANT: Navia, Manuel A.  
; APPLICANT: Ala, Paul J.  
; APPLICANT: Griffith, James P.  
; APPLICANT: Ali, Janid A.  
; APPLICANT: Faerman, Carlos H.  
; APPLICANT: Moe, Scott T.  
; APPLICANT: Magee, Andrew S.  
; APPLICANT: Connelly, Patrick R.  
; APPLICANT: Perola, Emanuele  
; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR  
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS  
ANTIBACTERIAL  
; TITLE OF INVENTION: DRUGS  
; FILE REFERENCE: 10283-014001  
; CURRENT APPLICATION NUMBER: US/10/186,886  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/301,676  
; PRIOR FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 44  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Mycobacterium avium  
US-10-186-886-44

Query Match 85.3%; Score 29; DB 14; Length 364;  
Best Local Similarity 71.4%; Pred. No. 4.3e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|||:|  
Db 262 FYDFATK 268

RESULT 11

US-10-282-122A-61746

; Sequence 61746, Application US/10282122A  
; Publication No. US20040029129A1

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61746
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-282-122A-61746
```

```
Query Match      85.3%; Score 29; DB 15; Length 369;
Best Local Similarity 71.4%; Pred. No. 4.3e+02;
Matches      5; Conservative      1; Mismatches      1; Indels      0; Gaps      0;
```

```
Qy      1 FYXFSTK 7
      || |:||
Db      266 FYDFATK 272
```

```
RESULT 12
US-10-186-886-42
```

```
; Sequence 42, Application US/10186886
; Publication No. US20030119061A1
; GENERAL INFORMATION:
; APPLICANT: Navia, Manuel A.
; APPLICANT: Ala, Paul J.
; APPLICANT: Griffith, James P.
; APPLICANT: Ali, Janid A.
; APPLICANT: Faerman, Carlos H.
; APPLICANT: Moe, Scott T.
; APPLICANT: Magee, Andrew S.
; APPLICANT: Connelly, Patrick R.
; APPLICANT: Perola, Emanuele
; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS
ANTIBACTERIAL
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 10283-014001
; CURRENT APPLICATION NUMBER: US/10/186,886
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/301,676
; PRIOR FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-186-886-42
```

```
Query Match          85.3%; Score 29; DB 14; Length 373;
Best Local Similarity 71.4%; Pred. No. 4.4e+02;
Matches      5; Conservative      1; Mismatches      1; Indels      0; Gaps      0;
```

```
Qy      1 FYXFSTK 7
      || |:||
Db      270 FYDFATK 276
```

```
RESULT 13
US-10-186-886-43
; Sequence 43, Application US/10186886
; Publication No. US20030119061A1
; GENERAL INFORMATION:
; APPLICANT: Navia, Manuel A.
; APPLICANT: Ala, Paul J.
; APPLICANT: Griffith, James P.
; APPLICANT: Ali, Janid A.
; APPLICANT: Faerman, Carlos H.
; APPLICANT: Moe, Scott T.
; APPLICANT: Magee, Andrew S.
; APPLICANT: Connelly, Patrick R.
; APPLICANT: Perola, Emanuele
; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS
ANTIBACTERIAL
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 10283-014001
```

; CURRENT APPLICATION NUMBER: US/10/186,886  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/301,676  
; PRIOR FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 43  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-186-886-43

Query Match 85.3%; Score 29; DB 14; Length 373;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYXFSTK 7  
|||:|  
Dt 270 FYDFATK 276

RESULT 14

US-10-186-886-45  
; Sequence 45, Application US/10186886  
; Publication No. US20030119061A1  
; GENERAL INFORMATION:  
; APPLICANT: Navia, Manuel A.  
; APPLICANT: Ala, Paul J.  
; APPLICANT: Griffith, James P.  
; APPLICANT: Ali, Janid A.  
; APPLICANT: Faerman, Carlos H.  
; APPLICANT: Moe, Scott T.  
; APPLICANT: Magee, Andrew S.  
; APPLICANT: Connelly, Patrick R.  
; APPLICANT: Perola, Emanuele  
; TITLE OF INVENTION: STRUCTURE-BASED DRUG DESIGN METHODS FOR  
; TITLE OF INVENTION: IDENTIFYING D-ALA-D-ALA LIGASE INHIBITORS AS  
ANTIBACTERIAL  
; TITLE OF INVENTION: DRUGS  
; FILE REFERENCE: 10283-014001  
; CURRENT APPLICATION NUMBER: US/10/186,886  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/301,676  
; PRIOR FILING DATE: 2001-06-28  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Mycobacterium smegmatis  
US-10-186-886-45

Query Match 85.3%; Score 29; DB 14; Length 373;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYXFSTK 7

Db            || |:||  
270 FYDFATK 276

RESULT 15

US-10-282-122A-64815  
; Sequence 64815, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64815  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-64815

Query Match                    85.3%;   Score 29;   DB 15;   Length 373;  
Best Local Similarity       71.4%;   Pred. No. 4.4e+02;  
Matches       5;   Conservative       1;   Mismatches       1;   Indels       0;   Gaps       0;

Qy 1 FYXFSTK 7  
|| |:  
Db 270 FYDFATK 276

Search completed: February 10, 2005, 16:41:31  
Job time : 42.9014 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2005, 15:38:08 ; Search time 10.8451 Seconds  
(without alignments)  
62.104 Million cell updates/sec

Title: US-10-067-484-4  
Perfect score: 34  
Sequence: 1 FYXFSTK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	30	88.2	116	2 S50449	hypothetical prote
2	30	88.2	201	2 G64013	hypothetical prote
3	29	85.3	94	2 H69748	hypothetical prote
4	29	85.3	222	2 T30423	hypothetical prote
5	29	85.3	258	2 T45991	hypothetical prote



6	29	85.3	373	2	B70673	probable ddlA - My
7	29	85.3	373	2	T34126	hypothetical prote
8	29	85.3	384	2	H87118	D-alanine-D-alanin
9	28	82.4	230	2	H70114	conserved hypothet
10	28	82.4	242	2	T16349	hypothetical prote
11	28	82.4	285	2	G72401	conserved hypothet
12	28	82.4	361	2	JH0417	cell adhesion mole
13	28	82.4	395	2	I77371	CD44R5 - human
14	28	82.4	426	2	JH0518	lymphocyte homing
15	28	82.4	464	2	A47442	olfactomedin precu
16	28	82.4	468	2	G70417	cytochrome oxidase
17	28	82.4	493	2	S13530	CD44E protein, epi
18	28	82.4	508	2	T22626	hypothetical prote
19	28	82.4	699	2	I37369	epican - human
20	28	82.4	742	2	A47195	lymphocyte homing
21	27	79.4	245	2	T33840	hypothetical prote
22	27	79.4	247	2	H64524	hypothetical prote
23	27	79.4	248	2	B97794	hypothetical prote
24	27	79.4	271	2	A95065	conserved hypothet
25	27	79.4	271	2	C97932	conserved hypothet
26	27	79.4	277	2	E75187	sugar abc transpor
27	27	79.4	277	2	D71220	probable sugar tra
28	27	79.4	297	2	A81381	hypothetical prote
29	27	79.4	304	2	T05587	hypothetical prote
30	27	79.4	328	2	A71981	DNA transformation
31	27	79.4	331	2	T20916	hypothetical prote
32	27	79.4	338	2	I40448	conserved hypothet
33	27	79.4	369	2	D90351	hypothetical prote
34	27	79.4	372	2	T25621	hypothetical prote
35	27	79.4	396	2	T39676	probable yeast cel
36	27	79.4	396	2	T24576	hypothetical prote
37	27	79.4	431	2	T20263	hypothetical prote
38	27	79.4	462	2	B88613	protein T27E9.5 [i
39	27	79.4	488	2	G81295	cytochrome-c oxida
40	27	79.4	510	2	I39930	replication protei
41	27	79.4	520	2	G88846	protein T12A7.2 [i
42	27	79.4	572	2	T47219	amino acid transpo
43	27	79.4	576	2	T25375	hypothetical prote
44	27	79.4	656	2	A72428	methyl-accepting c
45	27	79.4	656	2	E72379	methyl-accepting c

#### ALIGNMENTS

##### RESULT 1

S50449

hypothetical protein YEL010w - yeast (*Saccharomyces cerevisiae*)

C;Species: *Saccharomyces cerevisiae*

C;Date: 28-May-1993 #sequence\_revision 24-Feb-1995 #text\_change 09-Jul-2004

C;Accession: S50449

R;Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A;Description: *Saccharomyces cerevisiae* chromosome V cosmids 9871, 8199, 9867, 9495 and lambda clones 6693 and 5898.

A;Reference number: S50428

A;Accession: S50449

A;Molecule type: DNA  
 A;Residues: 1-116 <DIE>  
 A;Cross-references: UNIPROT:P40000; EMBL:U18530; NID:g602367; PID:g602377;  
 GSPDB:GN00005; MIPS:YEL010w  
 C;Genetics:  
 A;Gene: MIPS:YEL010w  
 A;Cross-references: SGD:S0000736  
 A;Map position: 5L  
 C;Superfamily: Saccharomyces hypothetical protein YEL010w

Query Match 88.2%; Score 30; DB 2; Length 116;  
 Best Local Similarity 71.4%; Pred. No. 7.1;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 :| |||  
 Db 31 YYSFSTK 37

# RESULT 2

G64013

hypothetical protein HI0787 - Haemophilus influenzae (strain Rd KW20)

C;Species: Haemophilus influenzae

C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004

C;Accession: G64013

R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.;  
 Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.;  
 McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;  
 Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.;  
 Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen,  
 D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.;  
 Geoghagen, N.S.M.

Science 269, 496-512, 1995

A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.;  
 Venter, J.C.

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.

A;Reference number: A64000; MUID:95350630; PMID:7542800

A;Accession: G64013

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-201 <TIGR>

A;Cross-references: UNIPROT:P44052; GB:U32762; GB:L42023; NID:g1573797;

PIDN:AAC22463.1; PID:g1573816; TIGR:HI0787

C;Superfamily: Haemophilus influenzae hypothetical protein HI0787

Query Match 88.2%; Score 30; DB 2; Length 201;  
 Best Local Similarity 71.4%; Pred. No. 12;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
 || |:|  
 Db 119 FYSFATK 125

# RESULT 3

H69748

hypothetical protein ybfE - *Bacillus subtilis*

C;Species: *Bacillus subtilis*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C;Accession: H69748

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: H69748

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-94 <KUN>

A;Cross-references: UNIPROT:O31445; GB:Z99105; GB:AL009126; NID:g2632457;

PIDN:CAB12012.1; PID:e1182170; PID:g2632504

A;Experimental source: strain 168

C;Genetics:

A;Gene: ybfE

Query Match \* 85.3%; Score 29; DB 2; Length 94;  
Best Local Similarity 71.4%; Pred. No. 9.8;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| |||:  
Db 24 FYFFSTR 30

RESULT 4

T30423

hypothetical protein ORF75 - *Lymantria dispar* nuclear polyhedrosis virus

N;Alternate names: Ld-bro-g

C;Species: *Lymantria dispar* nuclear polyhedrosis virus, LdMNPV

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T30423

R;Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohrmann, G.F.

Virology 253, 17-34, 1999

A;Title: Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*.

A;Reference number: Z20836; MUID:99124785; PMID:9887315

A;Accession: T30423

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-222 <KUZ>

A;Cross-references: UNIPROT:Q9YMQ2; EMBL:AF081810; PIDN:AAC70261.1

Query Match 85.3%; Score 29; DB 2; Length 222;  
Best Local Similarity 71.4%; Pred. No. 23;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
||| :|||  
Db 200 FYQFATK 206

RESULT 5

T45991

hypothetical protein F9D24.220 - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004

C;Accession: T45991

R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.

submitted to the Protein Sequence Database, January 2000

A;Reference number: Z23011

A;Accession: T45991

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-258 <DAN>

A;Cross-references: UNIPROT:Q9M2I5; EMBL:AL137081

A;Experimental source: cultivar Columbia; BAC clone F9D24

C;Genetics:

A;Map position: 3

A;Introns: 113/3

A;Note: F9D24.220

Query Match 85.3%; Score 29; DB 2; Length 258;  
Best Local Similarity 71.4%; Pred. No. 27;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
||| :|||  
Db 150 FYMYSTK 156

RESULT 6

B70673

probable *ddlA* - *Mycobacterium tuberculosis* (strain H37RV)

C;Species: *Mycobacterium tuberculosis*

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004

C;Accession: B70673

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence.

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70673

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-373 <COL>

A;Cross-references: UNIPROT:P95114; GB:Z83018; GB:AL123456; NID:g3261671;

PIDN:CAB05431.1; PID:g1694850

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: *ddlA*

C;Superfamily: D-alanine-D-alanine ligase

Query Match 85.3%; Score 29; DB 2; Length 373;  
Best Local Similarity 71.4%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FYXFSTK 7  
|||  
Db 270 FYDFATK 276

RESULT 7

T34126

hypothetical protein C26B2.8 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T34126

R;Du, Z.; Gattung, S.

submitted to the EMBL Data Library, November 1995

A;Description: The sequence of *C. elegans* cosmid C26B2.

A;Reference number: Z21480

A;Accession: T34126

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-373 <DUZ>

A;Cross-references: UNIPROT:Q18197; EMBL:U41559; PIDN:AAC24260.1; GSPDB:GN00022; CESP:C26B2.8

A;Experimental source: strain Bristol N2; clone C26B2

C;Genetics:

A;Gene: CESP:C26B2.8

A;Map position: 4  
A;Introns: 76/1; 116/3; 148/1; 201/1; 221/3; 267/1; 310/1  
C;Superfamily: Caenorhabditis elegans hypothetical protein C26B2.8

Query Match 85.3%; Score 29; DB 2; Length 373;  
Best Local Similarity 71.4%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| :|||  
Db 360 FYGYSTK 366

RESULT 8

H87118

D-alanine-D-alanine ligase A [imported] - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C;Accession: H87118

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Honore, N.; Ganier, T.; Churcher, C.; Harris, D.; Mungall, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Lacroix, C.; Maclean, J.; Moule, S.; Murphy, L.; Oliver, K.; Quail, M.A.; Rajandream, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, S.; Stevens, K.; Taylor, K.; Whitehead, S.; Woodward, J.R.; Barrell, B.G.

A;Title: Massive gene decay in the leprosy bacillus.

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: H87118

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-384 <STO>

A;Cross-references: UNIPROT:Q9CBS0; GB:AL450380; NID:g13093442; PIDN:CAC30631.1; GSPDB:GN00147

C;Genetics:

A;Gene: ddlA

C;Superfamily: D-alanine-D-alanine ligase

Query Match 85.3%; Score 29; DB 2; Length 384;  
Best Local Similarity 71.4%; Pred. No. 40;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| |:||  
Db 281 FYDFTTK 287

RESULT 9

H70114

conserved hypothetical protein BB0120 - Lyme disease spirochete

C;Species: Borrelia burgdorferi (Lyme disease spirochete)

C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004

C;Accession: H70114

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, N.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.

A;Reference number: A70100; MUID:98065943; PMID:9403685

A;Accession: H70114

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-230 <KLE>

A;Cross-references: UNIPROT:O51146; GB:AE001124; GB:AE000783; NID:g2688003;

PIDN:AAC66517.1; PID:g2688015; TIGR:BB0120

A;Experimental source: strain B31

C;Superfamily: conserved hypothetical protein YBR002c

Query Match 82.4%; Score 28; DB 2; Length 230;  
Best Local Similarity 71.4%; Pred. No. 41;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|| |||:  
Db 56 FYVFSTE 62

#### RESULT 10

T16349

hypothetical protein F42G9.9 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C;Accession: T16349

R;Taich, A.

submitted to the EMBL Data Library, March 1996

A;Description: The sequence of *C. elegans* cosmid F42G9.

A;Reference number: Z18498

A;Accession: T16349

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-242 <TAI>

A;Cross-references: EMBL:U00051; NID:g1216305; PID:g1216307; PIDN:AAA91353.1;

CESP:F42G9.9

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:F42G9.9

A;Introns: 58/3; 141/3

Query Match 82.4%; Score 28; DB 2; Length 242;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFSTK 7  
|: |||  
Db 185 FFRFSTK 191

RESULT 11

G72401

conserved hypothetical protein - *Thermotoga maritima* (strain MSB8)

C;Species: *Thermotoga maritima*

C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004

C;Accession: G72401

R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Nature 399, 323-329, 1999

A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*.

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: G72401

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-285 <ARN>

A;Cross-references: UNIPROT:Q9WY71; GB:AE001707; GB:AE000512; NID:g4980720;

PIDN:AAD35320.1; PID:g4980727; TIGR:TM0229

A;Experimental source: strain MSB8

C;Genetics:

A;Gene: TM0229

C;Superfamily: *Methanobacterium thermoautotrophicum* conserved hypothetical protein MTH1382

Query Match 82.4%; Score 28; DB 2; Length 285;  
Best Local Similarity 83.3%; Pred. No. 51;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FYXFST 6  
|| |||  
Db 39 FYSFST 44

RESULT 12

JH0417

cell adhesion molecule CD44 - human

C;Species: *Homo sapiens* (man)

C;Date: 23-Nov-1991 #sequence\_revision 23-Nov-1991 #text\_change 09-Jul-2004

C;Accession: JH0417; A32376; G02251; A32377

R;Harn, H.J.; Isola, N.; Cooper, D.L.

Biochem. Biophys. Res. Commun. 178, 1127-1134, 1991

A;Title: The multispecific cell adhesion molecule CD44 is represented in reticulocyte cDNA.

A;Reference number: JH0417; MUID:91337049; PMID:1840487

A;Accession: JH0417

A;Molecule type: mRNA

A;Residues: 1-361 <HAR>

A;Cross-references: UNIPROT:Q92493; GB:M59040; NID:g180129; PIDN:AAA51950.1; PID:g180130

A;Experimental source: reticulocyte